BUCKINGHAM

RESEARCH
FOR
TEACHERS

RESEARCH FOR TEACHERS

BY

BURDETTE ROSS BUCKINGHAM

DIRECTOR OF THE BUREAU OF EDUCATIONAL RESEARCH AND
PROFESSOR OF EDUCATION, THE OHIO STATE UNIVERSITY

SILVER, BURDETT AND COMPANY
W YORK NEWARK BOSTON CHICAGO SAN FRANCISCO

~~371~~
~~B85~~

COPYRIGHT, 1926, BY
SILVER, BURDETT AND COMPANY

Printed in the United States of America

LB
1027
.B8

PREFACE

I HAVE attempted in this book to bring together in an accessible way some of the results of research work which I think the classroom teacher will find most helpful. In doing so I have pretty rigidly excluded from consideration many brilliant analyses and significant conclusions in the field of school administration. Matters of organization in school work, financial questions, the development of the curriculum, and the interesting results of child accountancy—except as they apply in the classroom—have been disregarded. Not that these questions are outside the proper interest of the teacher. I hold that the teacher *is* properly interested in such matters, and that in the schools of the near future teachers will more frequently and more positively participate in administrative matters because they will be more competent to do so.

I have, however, excluded administrative and supervisory concerns from the field of this book largely because there are so many things I want to say to classroom teachers with reference to classroom work that there seems to be no room for anything else. Even as it is, I feel no small uneasiness regarding the matters that are not included in my selection of topics. Almost any person well-informed as to the development of educational research could write another book for teachers which should have much the same objective as

20518

this one, but which would contain entirely different matter. Sins of omission, therefore, I am as keenly conscious of as the most penetrating critic.

The purpose of the book is twofold. One purpose, as has been intimated, is to show the teacher some of the things he can use in his work—things which have been developed, not merely by appeals to principles, but primarily by methods of experimentation. The other purpose—and it seems to me to be quite as important—is to show that the teacher has opportunities for research which, if seized, will not only powerfully and rapidly develop the technique of teaching but will also react to vitalize and dignify the work of the individual teacher. It is my firm belief that the emancipation and professionalizing of the teacher's calling rests far more upon the originality, insight, and expertness which the teacher evinces than upon any considerations having to do with salary, tenure, or legal status. Society cannot be compelled to respect anybody or anything. The surest way to win respect is to be respectable. If, as has often been observed, there is no such general belief in the expertness of teachers as there is concerning the expertness of those engaged in other occupations, it is worth considering whether or not the public is right about it—whether the teacher really has, to the extent that the professional or craft worker has, a special competency peculiar to his calling. In my judgment, nothing would so effectively obtain for the teaching body the possession of professional expertness and the reputation for having it as the open-eyed, open-minded, scientific spirit of inquiry.

In what I have written I have sought to appeal especially to the teacher in service. With my audience so definitely in mind I have ventured on some occasions to employ the form of direct address. I have used "you" and "your" rather freely. Moreover, for the sake of coming closer to my audience I have made another departure from literary convention. When I speak in my own person I do not resort to the editorial "we" or to the passive voice. I use "I" and "my" and "mine" when these words convey my meaning. I hope I shall not give offense through my failure to hide behind a circumlocution which in the light of the purpose of this book would seem to me to be a species of affectation.

In preparing this book for publication I have been indebted to Miss Josephine MacLatchy, Editorial Assistant in the Bureau of Educational Research of Ohio State University, for help in copy editing the manuscript, to Dr. Edgar A. Doll for valuable criticism, especially with reference to the first chapter, and to my colleague, Dr. Ward G. Reeder, and Robert K. Toaz, Superintendent of Schools, Huntington, N. Y., for reading the proof and for numerous helpful suggestions.

B. R. BUCKINGHAM.

Ohio State University.

CONTENTS

CHAPTER PAGE

I. SOME THINGS WE KNOW ABOUT LEARNING 1

II. WHAT STATISTICS DOES THE MODERN TEACHER NEED? 73

III. BETTER TESTING: INTELLIGENCE TESTS 133

IV. BETTER TESTING: EDUCATIONAL TESTS 167

V. COMBINING INTELLIGENCE AND EDUCATIONAL TESTS . 198

VI. NEW-TYPE EXAMINATIONS 210

VII. GROUPING AND CLASSIFYING PUPILS 258

VIII. ERROR AND FAILURE: THEIR EDUCATIONAL MEANING . 295

IX. REACHING THE INDIVIDUAL 330

X. THE TEACHER AS A RESEARCH WORKER 369

INDEX 381

RESEARCH FOR TEACHERS

CHAPTER I

SOME THINGS WE KNOW ABOUT LEARNING

No matter how we define education there will be substantial agreement that it is a process of adjustment. It seeks to effect certain changes in individuals. Born with certain instinctive tendencies, man is ill fitted to participate in social life until his inherited instincts and capacities have been altered into socially useful habits, interests, and powers. Perhaps the most characteristic of all his inherited capacities is his capacity for growth. It is this which distinguishes him from the brute creation. In virtue of this well-nigh limitless possibility for growth, he has created for his habitation a world so complex that adjustment to it has become a long and elaborate process—a process requiring an extended period of infancy and dependence.

THE TEACHER AS THE AGENT OF SOCIETY

Moreover, it has been found that the adjustment of the immature members of society to the life we call civilized cannot be left to the unguided forces of immediate environment. Influences from far and near must be assembled, and the experience of the race must be drawn upon. The here and now are insufficient. The growing organism, in order to qualify in highly developed modern life, must be subjected to impressions beyond the scope of its senses. It must live in imagination with people of other lands. It must draw from the past the best that has been said and done. It must come into the possession of its inheritance.

Society must put forth great effort in order to make this possible. The task is tremendous and at the same time inescapable. Moreover, it cannot be left to the good will of parents. Indeed, with the development of civilization, the task has become so exacting that few if any members of society possess sufficient wisdom to perform it in its entirety. Accordingly, an institution has been created and a special class of people has come upon the scene—a class whose business it is to collect, organize, condense, and present the most appropriate materials for use in training the young. This institution is the school and the special class is the teaching profession.

This profession, therefore, is obviously interested primarily in two things—the material which we call the curriculum and the use of it for instructional purposes. In the specialization of activities which the teaching profession, like every other calling, has exhibited, the curriculum has come to be the concern of one group of specialists, and its use or application has become the task of a different group. I am not concerned at the moment with the advisability of this division of labor. It is enough to point out that in general superintendents are responsible for the curriculum and teachers for putting it into effect.

TEACHING PARALLELS LEARNING

But since it is in the child and not in something or somebody else that the capacity for growth inheres, it is clear that the only way this organized material which we call the curriculum can be put into effect is in accordance with the way in which growth takes place. The wisest method of teaching and the most effective is the one which conforms most closely to the way the child learns. Thus it comes about that the good teaching method parallels the good learning method; and that the test of a method is found not in its logical completeness, but in its psychological conformity to the nature of the child.

You will readily see that I am making growth and learning somewhat synonymous. Of course, this is not strictly true, yet in mental life growing and learning are very closely related. Even though the two are by no means identical, they are much more nearly so in regard to school work than they are in regard to life outside of school. Accordingly, for our purpose we shall make no serious mistake if we take the stand, first, that growth is a fundamental objective in education; second, that this requires in a civilized environment a prolonged period of training; third, that in order to effect this training—in order to realize this growth—society consciously sets about the task, thereby creating as its agents the school and the teaching class; fourth, that the capacity of the child for growth, his ability to take on experience, his receptivity to the efforts of the school and the teacher is his capacity to learn.

PASSIVE AND ACTIVE LEARNING

Learning is both passive and active. About us are the phenomena of life impinging upon our senses and giving rise to the experience we call sensation. One of our human attributes is an experiencing nature—an attribute which we share, to be sure, with the beasts of the field. But we possess this experiencing nature in an especially high degree and, whether we will or not, these phenomena which surround us—these forms and colors, these sounds, tastes, and smells—leave in us traces by virtue of which we not only apprehend them at the moment, but retain them afterwards. The effect of these influences is registered without effort on our part. From this point of view, therefore, the taking on of experience—the thing we call learning—is passive.

We shall not have much to do in this chapter, nor indeed in this book, with experimental results in passive learning. This field, so far as it relates to the school, has been little considered—certainly less than it ought to be. To be sure, we have that relatively passive type of learning which Meumann distinguishes as "observational." In this observational field a very great amount of investigation has taken place under the general heading of "testimony." The results, although profoundly important in their bearing upon the science of psychology, need not especially concern us here.

There is, however, a sense in which passive learning is exceedingly important in school work. The whole question of the right surroundings of the child while at school belongs under this heading. As I have said, experimental results here are lacking, but we should not on that account fail to realize the undoubted significance of the passively experiencing nature upon the development of mental and emotional life. It seems likely that such matters as attitude, taste, *esprit de corps,* and personality are profoundly affected by the incidental influences with which children are surrounded.

I cannot avoid mentioning in this connection the attitude of the ancient Greeks, although I fancy some will smile in a superior way and point out that the ideas entertained in Athens twenty-four hundred years ago are very far removed from the presumed rigor of modern educational research. But they *got results;* and that idea is modern enough—as modern as "efficiency." As to these results, Shelley had this to say: "The human form and the human mind attained to a perfection in Greece which has impressed its image on those faultless productions whose very fragments are the despair of modern art, and has propagated impulses which cannot cease, through a thousand channels of manifest and imperceptible operation, to ennoble and delight mankind." If this is a fact, it is as worthy of consideration as a measure of intelligence or a standard in arithmetic.

It is conceded that Plato in his *Republic* set up a plan of education essentially like that which was going on about him. If, therefore, we desire to find out how this perfection of the human form and mind to which Shelley referred was attained in ancient Greece, we can do no better than read the *Republic.* When we do this, we find that there is little evidence of the conscious effort and drive of modern education. The fundamental demand of Plato was that the soul should be surrounded with that which should strengthen and beautify it. "Ought we not," he asks, "to seek out artists . . . who are gifted to discern the true nature of the beautiful and graceful; then will our youth dwell in a land of health, amid fair sights and sounds, and receive the good in everything; and beauty, the effluence of fair works, shall flow into the eye and ear, like a health-giving breeze from a purer region, and insensibly draw the soul from earliest years into likeness and sympathy with the beauty of reason."[1]

Of course, this idea of the importance of an educative environment was not new with Plato. It underlay a great deal of what was best in the practice of his own and of previous times. The necessity for it, now as well as then, cannot reasonably be denied. A great deal of the influence with which we seek to surround pupils in our best schools is merely a part of what in a broad sense constitutes an atmosphere which we believe to be conducive to the proper development of human life. Concretely it has to do with such matters as fine school buildings and grounds, good pictures and statuary in schoolrooms, lectures, music, drama, and all things curricular and extra-curricular which may be brought to bear advantageously upon the experiencing nature—Plato called it the "soul"—of children.

On the other hand, so exacting have become the demands of life that passive learning is entirely insufficient. Acting under the spur of instinctive tendencies and whipped into activity by social sanctions and penalties, the human being puts forth active effort to learn more and better than he would learn in a passive or merely receptive way. At this point the school and the teacher come prominently upon the scene. A large part of the business of the teacher is to facilitate conscious, active learning. He does this in general by two procedures: first, by condensing, organizing, and adapting masses of experience so that they may be quickly and effectively apprehended; and second, by devising a superior technique for introducing or presenting this adapted material to the learner.

---

[1] *Republic,* Book III: 401, C-D.

I shall devote the remainder of this chapter to a consideration of the way in which, for the most part in an active sense, the child learns. I shall do this in order to indicate some of the ways in which the teacher may perform successfully this second of the teaching functions, namely, the utilization of an effective teaching method.

THREE FUNDAMENTAL LAWS

Learning takes place only when certain fundamental conditions exist. In the first place, the child cannot learn unless he is ready to learn. You cannot teach long division to a child in the first grade. You cannot interest him in abstract reasoning. You cannot teach anybody anything until he has some preparation, some point of departure upon which the teaching may be based. In other words, pupils must be ready for our ministrations, and this fact has been referred to by Thorndike as the *Law of Readiness.*

In the second place, no more fundamental precept has ever been uttered in connection with learning than that *we learn to do by doing.* Every act we perform tends, other things being equal, to be performed again more easily.

The highly skilled person is a person who has done the thing in question many times. On the other hand, *we unlearn by not doing*. Even a habit may be lost through disuse. Many examples of the effect of use and disuse upon learning could be given. It seems hardly necessary, however, to labor with the point. Column addition is not learned without adding many columns. Sewing is not learned without sewing, nor football without playing football. Repetition, we are agreed, is essential in learning. We are therefore prepared to assent to the second of these fundamental conditions—namely, the *Law of Exercise.*

But acts are repeated which give pleasure and acts are discontinued which give pain. The law of exercise, about which I have just spoken, comes into play largely according to the effect the act in question has upon the organism. Accordingly, we have, in the third place, a law called the *Law of Effect.* If a child, or an adult for that matter, finds an experience pleasing, he tends to repeat it or to prolong it, thus securing the benefit, in his learning, of the law of exercise. On the other hand, if an experience proves to be unpleasant, he does not repeat it or he discontinues it. Again the law of exercise is put into effect, this time in a negative way; but this negative learning is no less important than positive learning. It may be as necessary for the child to dread the fire through the unpleasant experience of being burned as it is for him to learn to walk or to talk, to read or to extract square root. These positive acquisitions are facilitated by the satisfaction arising from the experience connected with them. Sometimes the satisfaction is wholly relative and may be artificially created. For example, the experience of extracting square root may not be inherently satisfying, but matters may be so arranged that it becomes more satisfying than the experience derived from failure.

I have thus referred to three fundamental laws, namely, those of *readiness*, *exercise*, and *effect*. It is through the action of these laws that the child learns. It is through their operation that society, partly by the action of its agents, the teachers, modifies or adapts, strengthens or weakens, preserves or thwarts original tendencies. It is through the operation of these laws that old and harmful instinctive responses to situations are suppressed, while new and beneficial responses to the same situations are developed. Curiosity, for example, from being a source of mischief becomes the means by which the inquiring child is led into his inheritance.

The law of readiness. These laws, however, are too general to be of much use in the classroom. A teacher does not ordinarily say, "I shall now put the law of readiness into effect;" or, "I shall now exhibit the operation of the law of exercise"; or, "This method which I am about to pursue brings the law of effect into play." In order that these laws may be put to use, they must be analyzed down to working units. Thus we shall find that applying the law of readiness brings us first to a consideration of the native abilities of the pupil.

We know so much more today than we used to know about the conditioning influence of native ability. We have learned that some children are not ready for a given learning task as soon as others and that a few never become ready for any but the simplest tasks. A child who has the mentality of a typical child of six—that is, whose *mental age* is six—can profitably begin the first grade of an American school—unless, of course, he is handicapped otherwise than mentally. He can learn to read as children are taught reading in that situation. On the other hand, if his mental age is less than five and one-half he cannot be taught reading as it is taught in our first grade, unless other conditions are especially favorable. In other words, he is not ready. Similar statements may be made with reference to the readiness, so far as mentality is concerned, of pupils in any grade.

Do you know whether or not each of your pupils is ready for the work you are going to require? If they have not had intelligence tests you should attempt to secure a set of such tests; and, after studying them carefully, you should test your pupils with them. The results will be useful to you in many ways. Not the least of these will be in the knowledge you will get of the mental readiness of your pupils for the tasks the school is expecting them to perform.

Of course, there are a great many other senses in which children may be ready or not ready for their tasks. Most of these you will have to estimate just as, a few years ago before the tests came into existence, you would have had to estimate mentality. But mentality is the most important factor and if to it you can add a good judgment as to capacity for effort, you will have a very serviceable representation of each pupil's native abilities so far as they affect his readiness to learn.

But there is much more in the application of the law of readiness than that which relates to the native abilities of pupils. A pupil may be ready in a general sense and be far from ready at the moment. He may have developed an attitude toward the task in hand which may prevent his learning as long as the attitude lasts. The prevention of unfavorable attitudes is, therefore, an item under the general heading of the law of readiness. Much the same thought may be expressed positively as the provision of proper physical and mental influences.

There are a number of specific things for the teachers to do to secure a readiness for the work in hand. One of them is seeing that the child understands exactly what he is to do. Another is preparing the learner for his task by explaining its difficulties. Still another is the bringing up of related past experience.

A great deal has been done by the teacher, however, when he has provided conditions favorable to the learning act. Professor Book, in his excellent analysis of what the teacher can do to expedite learning, gives special prominence to this idea.[2] It should be understood, however, that this provision of favorable conditions refers not only to the early stages of the learning, but also, as we shall see later, to its entire course. There are critical periods in any learning process—periods in which progress is not only arrested, but threatened with discontinuance. The rôle of the teacher at these junctures, no less than at the beginning, is that of supplying favorable conditions.

---

[2] Book, William F. "The Rôle of the Teacher in the Most Expeditious and Economical Learning," *Journal of Educational Psychology*, 1: 183-99 (April, 1910).

The law of effect. But undoubtedly an important consideration in supplying favorable learning conditions is the emotional response which the act of learning produces. Thorndike, whose simplicity in the treatment of the laws of learning seems to me to commend his doctrine to practical workers, asserts in substance that learning is connecting and that man is the great learner primarily because he forms so many connections.[3] Another name for connection is, of course, association; and, other things being equal, the connections or associations which are accompanied by a feeling of satisfaction or pleasure will be firmly established. Learning will thus be favored; and throughout its course the teacher's part will be to see that the learning situation is accompanied by a pleasurable "feeling tone." The law of effect, which asserts the importance in learning of this feeling tone, is thus operative at all stages of the learning process. It is intimately connected with interest, though it is more fundamental. How to sustain it in the classroom becomes a basic question in the art of teaching and of schoolroom management. The study of what interests children has been experimentally attacked in only a few instances.[4] It deserves considerably more attention from the psychologists. Moreover, good teachers in putting the law of effect into practice—as all good teachers do—have really, though often unconsciously, analyzed the operations of this law. Both the teachers and the psychologists, however, have failed lamentably to put the analysis into usable form and to transmit it to others.

---

[3] Thorndike, E. L. *Educational Psychology, Briefer Course.* New York: Teachers College, Columbia University, 1919. p. 173.

[4] See especially in this connection Jordan, A. M. *Children's Interests in Reading.* New York: Teachers College, Columbia University, 1921; and Dunn, Fanny W. *Interest Factors in Primary Reading Material.* New York: Teachers College, Columbia University, 1921.

The law of exercise. In the case of the law of exercise analysis is likewise needed before the teacher can go to work intelligently. Merely to know that the occurrence of a reaction tends to make its subsequent occurrence easier is to possess little guidance in the matter of adapting one's teaching method to the learning process. So exceedingly important is this question of analysis and so broadly does the law of exercise cover the conditions of the learning process that I shall stop here to consider some of the ramifications of this doctrine. What we know about learning—what the psychologists and experimentalists have actually determined—falls for the most part under the heading of the law of exercise.

If we say that an action once performed tends to be repeated or that an idea once entertained tends to be reproduced, we do nothing more than assert the ability of the human organism to remember. This is the capacity which we call memory. It is possessed not only by human beings but also by animals. It has been extensively studied, and a great deal of what we know about learning has been acquired through these studies of memory.

PERIODS OF PRACTICE

Although a reaction by its occurrence tends to reproduce itself, most learning of the formal type requires frequent repetition; and one of the questions on which the experimentalists have thrown light is that of the distribution of this repetition. In the school, repetition is called *practice* or *drill.* The findings of the investigators, therefore, regarding the distribution of practice are evidently of no small interest to teachers. One question in this connection is, "How long should practice continue at one time?"

It is pretty well established that if one is going to devote a given amount of time to a learning task, short work periods will be more beneficial than long ones. For example, Starch found that in learning to associate numbers with letters—namely, learning to write numbers for letters according to a code—"it is more economical within limits to shorten the periods of work and to distribute them correspondingly over a given period of time."[5] His evidence was particularly impressive, and it has been confirmed by other experimenters working with different materials and, of course, with different learners. In Starch's experiment four different groups of persons worked, each for 120 minutes. The first group worked ten minutes twice a day for six days. The second worked twenty minutes once a day for six days. The third worked forty minutes at a sitting on three alternate days. Finally, the fourth group worked 120 minutes continuously. Observe then that the periods of work at a sitting were ten, twenty, forty, and one hundred twenty minutes. Now, the efficiency of the work was precisely in the order I have just mentioned, the most efficient being done in the ten-minute period and the least efficient in the 120-minute period. There was not a great difference between the success of the ten-minute group and that of the twenty-minute group, but the forty-minute group was very much inferior and the 120-minute group was still more so.

---

[5] Starch, Daniel. "Periods of Work in Learning," *Journal of Educational Psychology*, 3:209-13 (April, 1912).

Application to arithmetic. In general, then, short periods of learning, practice, or drill yield better results than long periods. It will be considerably more effective to drill on arithmetic for ten-minute periods than for one-fourth as many periods each forty minutes long. Of course, I am now drawing an inference which is somewhat beyond the original facts. The inference is that from the type of material which Starch used, I can make an assertion about drill in arithmetic. If there were no corroborating evidence, all I should be justified in saying would be that the results Starch obtained raised a probability that similar results would be secured in arithmetic—the probability being the greater the more closely his experimental material and method resembled the material about which I made the inference. It happens, however, that in this case direct corroborative evidence is available. Kirby has furnished it in his experiments on drilling children in arithmetic. He finds precisely what Starch found, namely, that a given amount of time can be most effectively used in short, frequent periods.[6]

The periods can be made too short. At this point, however, it is well to insert a caution. You cannot indefinitely shorten the periods. It takes time to make the necessary arrangements—to collect or to distribute material, to put away the books or other equipment used in the former work.

---

[6] Kirby, Thomas J. *Practice in the Case of School Children.* New York: Teachers College, Columbia University, 1913, pp. 54-55, 60-63.

But the most important reason for setting a limit to the shortness of the period has to do with the fact that every learner requires some time in which to adjust himself. A number of experiments have been performed, the object of which has been to show the way in which people begin a learning task. Some of them attack the job with a rush; and these are likely to tire rather quickly. For them the short period is particularly advantageous; and in extreme cases the period may properly be very short. Other learners take considerable time in adjusting themselves to the task in hand. They reach their best pace only after a relatively long interval. Persons of this type usually show more endurance. In the experimental work this fact is shown by their continued freedom from error over a relatively long period. Between these two extremes lie all gradations as to the length of time required to become adjusted.

The teacher has here an excellent opportunity for research work of his own. Evidently if he is going to take account of individual differences this question of the best length of learning period *for each pupil* will yield large returns in repayment for the study of the pupil's methods of learning. In this, as in so many other respects, there is a difference between the teacher who is a mere practitioner and the teacher who is a real expert.

The "best" length of period. As to the proper length of drill period, some differences no doubt will exist among children of different grades and also among children studying different subjects. The type of learning required—whether it be verbal, manual, or logical—will also have bearing on the question; and as I have already indicated, the individual differences among children must be considered in making any accurate adjustment. Some of these questions have been investigated and if I were writing a book instead of a chapter on "Some Things We Know About Learning" I should enter into these matters.

Under present circumstances, however, I shall have to be content with pointing out that the question of the proper length of the practice period, after we know that in general it should be short and frequently repeated, resolves itself into a compromise for the group of pupils constituting a class—a compromise based on the recognition of individual differences and of the average condition of the class. No information which you can derive from the experiments of others—valuable as that information is—will vie in importance or applicability with the results of your own investigations with the pupils of your own class. Try Starch's experiment. Test your pupils, say, for column addition. Then let some of them practice for short periods and others for longer periods. Two things are essential: (1) the ability of each group of children in column addition must be the same at the beginning; and (2) the total practice time must be the same for all pupils. It would be well, too, to space out the periods, as Starch did, so that all the groups will finish on the same day. Finally give at the end of the series of practice periods another test in column addition as nearly like the first one as possible. Indeed, you may give the *same one*. Note now how much greater each child's score on the final test is than his score at the beginning. Get averages for each practice group. Without telling the pupils you are going to test them again, give them a similar test in column addition a month later. This is to see whether there is any difference between the groups in retention.

There is nothing formidable about this experiment, and you can do it without upsetting your regular work. Try it. The question of length of work periods will take on new meaning for you.

WHY SHORT PERIODS ARE BETTER

If we ask why short periods are economical in learning, at least three facts may be presented, each of which has behind it the support of substantial evidence.

First, when the learning exercises are short and frequent there are periods of rest between them during which the associations become settled and perfected. Indeed, a very important part of learning goes on during these periods of rest. This is what led James to say that we learn to skate in summer and to swim in winter. The fundamental idea is enunciated in Jost's law or *the law of the age of associations*. This law may be stated as follows: "Of two associations which are of equal strength but of different ages the older receives the greater intensification from new repetitions." [7] Now, in Starch's experiment in learning numbers for letters, or in Kirby's experiment in arithmetic drill, if a person has had ten minutes of practice followed by a day without practice and if he enters upon a second ten-minute practice period the following day, it is clear that at that time his associations derived from the first ten minutes of practice are much older than the associations derived from ten minutes of practice by the person who without a break at the close of the first ten minutes goes on to practice for a longer period. If these older associations have not been too much weakened by the onset of forgetting—as in general they appear not to be—Jost's law asserts that they are more readily refreshed and reënforced than the younger associations which have been more recently acquired.

---

[7] For example, father and son may do equally well on an algebra test. Their associations are at present of "equal strength." If both father and son now study algebra each to the same extent ("new repetitions"), the father will surpass the son on a second test.

Second, shorter periods are better than longer periods, because in longer periods feelings of fatigue play a greater part. Ever since the first scientific experiments on memory by Ebbinghaus back in 1880, it has been observed that after a certain period, differing as usual with different individuals, additional study of the material to be memorized has little effect. This is no doubt partly explained by the onset of boredom, distraction, and perhaps of real fatigue.

Third, and in consequence of the above, short periods of learning are likely to be more effective, because during such periods a higher degree of concentration of attention is possible. Common experience, as well as experimental evidence, teaches us that mere repetition of material to be memorized or learned is of little value. It is *attentive* repetition that counts. Accordingly, time spent in learning will be most highly rewarded when it is characterized by the sharpest concentration of attention.

LEARNING WHICH INVOLVES LITTLE REPETITION

Observational. In the act of learning, as I have already stated, repetition is an important consideration. It is true, however, that there are types of learning in which repetition plays a subordinate part. One of these is the activity which goes on when a person merely notes what presents itself to his senses. It is exemplified when a child looks at a picture and tells what he sees. This sort of activity assumes large importance in life. The ability to pass through a street and remember something about it, or the ability to meet people and remember their faces, to say nothing of their names, is an ability in observational learning which is of great practical utility. I am sure that if our school work were as closely related to life in fact as it is in theory we should frequently place children in positions which favor careful observational noting.

The reason that repetition is of minor importance in so-called observational learning is that the conditions of such learning do not ordinarily permit it to take place. For the most part the experiences involved are met but once. It is not to be supposed that repetition would not be important if it were applicable, but rather that from the nature of the case the recurrence of the observation is precluded. No one doubts that a second or third presentation of a picture will permit a more accurate recall of its details or that meeting the same person several times will facilitate the recognition of him after an interval. All that we are now interested in showing is that as experience comes to us and to our pupils there is much that comes but once. The requirements of life, therefore, make necessary the learning of many things at a single presentation.

Logical. But observational learning is not the only type in which repetition is or may be eliminated. In logical learning, especially when the material is highly organized, richness of association may suffice to insure relatively complete learning. Much, however, depends upon the extent and variety of the association with previous experiences—which is the same as saying that much depends upon the extent to which the matter in hand is understood. If something that I write in this chapter relates closely to your experience so that you thoroughly apprehend it, you will confidently rely upon your understanding of the matter to assure you of its possession. Here repetition is not reduced to a minimum because it does not apply—as in observational learning—but rather because learning is sufficient without it. Again, however, repetition may help. Much depends upon your purpose and the degree to which you wish to learn. If it is your purpose to tell someone else what you learned from a certain passage of reading matter, you will undoubtedly make a better report if you re-read the passage.

I have made this excursion because I wanted you to see that there are types of learning which, under ordinary circumstances, involve little repetition. Your pupils should be encouraged to learn many things without repetition—by observation, in fact—because there are thousands of occasions in which they will have to order their lives on the basis of this type of learning. On the other hand, you should seek where you can to make your teaching so meaningful, so related to past experiences, and so interrelated in its parts that repetition may for ordinary purposes be dispensed with.

Try your class on observation. Show them a picture. Let them take a good look at it for, say, thirty seconds.

Then let them write answers to a series of questions—*e.g.* Was there a policeman on duty when the collision occurred? What time was it (a clock tower in the distance shows)? What was the name of the street on which the trolley car was running? What was the license number of the automobile? What kind of weather was it?

Let them have an exercise like this, perhaps every day for a week. See if you can devise a way to measure their improvement in this type of observational learning.

HOW MUCH TO REPEAT: THE WHOLE AND PART METHODS

But most types of learning depend to a large degree upon repetition. The study of this process of repetition is, therefore, important. Like most activities, it can be well done or poorly done.

What is meant by the whole and part methods. For example, it appears to make a great deal of difference how *much* material one repeats at a time. If one has a rather extensive learning task to perform, it has been found to be more advantageous to distribute attention over the whole material than to give separate attention to its parts. Doubtless you have heard of the *whole* and *part* methods of memorizing. These words—"whole" and "part"—are convenient terms to describe extreme methods which, as a matter of fact, few people use. On the one hand, the whole method without qualification requires that a memory task be surveyed repeatedly from beginning to end until it is "learned" in its entirety. (My deprecatory quotation marks around the word learned are intended to suggest its ambiguity. In any given task the degree of learning required is part of the task.

To say that a thing is learned is to leave matters in uncertainty until we define what is intended.) On the other hand, the part method of learning supposes that the material is divided into small—how small is undetermined—parts, all of which are first learned separately. The parts are then put together by an additional effort of learning so that the whole may be reproduced.

Between these well-nigh impossible extremes there is every gradation. We may, however, for practical purposes, consider learning to be characterized by the whole method when the movement of attention is essentially progressive from the beginning to the end of the material. This loose conception permits the material to be divided into parts and pauses to be made at the conclusion of each part, but it does not permit the learner to turn back to the beginning of one of the parts after he has finished it. This method is essentially the one suggested by Meumann as a "mediating procedure." [8]

As to the part method, the essential idea is that of discontinuity. One not only divides what one has to learn but also learns each division as a separate act. The only permissible variation is in the way the parts are put together. As I have said, in the extreme type no parts are brought together until all of them have been separately learned. According to more usual methods, the parts are at least partially linked together as they are learned. For example, one may learn the first two parts separately and then connect them as soon as the second part is learned. The third part will then be learned separately and immediately added to the first two; the fourth will be learned and added to the first three; and so on. This is what Pechstein calls the "progressive part method," and he finds it very superior, although the material he used is not the kind one has to deal with in school. Another method of putting the parts together is to repeat the first part as an introduction to the second throughout the entire learning of the second part. Then the first and second parts together constitute the introduction to the third part and are repeated with it while it is being learned. Then the first, second, and third parts are repeated as the introduction to the fourth part; and so on. This would be intolerable in a long learning task.

---

[8] Meumann, E. *Psychology of Learning* (translated by Baird). New York: D. Appleton and Co., 1913. p. 253.

The evidence in general. The preponderance of evidence is to the effect that a "whole" method of learning is better than a "part" method. For example, summarizing the experimental evidence on the question up to about 1911, Meumann says, "We may therefore regard the following general result as established: for adults and children it is more advantageous and it is psychologically and pedagogically more appropriate to learn every sort of material as a whole than it is to break it up into parts."

Subsequent investigations have not brought this statement into serious question. If we consider Meumann's summary as referring to the extreme whole and part methods, the only real dissenting voice, so far as I have learned, is that of Winch, whose results for children of about the ages of ten to twelve show in the learning of poetry a somewhat greater advantage for the part method.[9]

Measuring learning: immediate memory. When I speak of a method of learning as "better," I should define what I mean, and this can be most satisfactorily done in numerical terms. That method of learning is better which gets the same things learned with less effort or gets more things learned with the same effort.[10] We can apply the first—getting things learned with less effort—by defining "things learned" and measuring the effort. Now a thing is learned, common sense tells us, when it can be reproduced without error; and the experimentalists have often used this definition. The effort, then, required to reach "a first errorless reproduction" may be measured by the time it takes to attain this result. If, for example, a child studies a stanza of Byron's *Childe Harold* and if it takes him eight minutes to reach the point where he can repeat it without a mistake, the measure of his learning effort is eight minutes. Another measure of his effort might be the number of readings or repetitions of the stanza required before it could be recited without error.

According to either of these ways of measuring—namely, taking the time or counting the repetitions—the number of errors is ruled out by requiring a response that is free from error. But we may render the time constant and measure the number of errors—as some do—by giving each child the same task and the same amount of time for it. The measure of the learning may then be the amount of the original material which is correctly reproduced—or, inversely, the number of errors made in reproduction. This measurement, however, of the work of learning by counting the errors or the number of correct items is by no means easy. It involves breaking the material up into items which shall have substantially equal difficulty for the learner. This is an almost insuperable difficulty with any but the simplest material; and even in such material as words to be spelled, dates, or vocabularies in a foreign language to be learned, the differences in difficulty are by no means small.

---

[9] Winch, W. H. "Should Poems Be Learnt by School Children as 'Wholes' or as 'Parts'?" *British Journal of Psychology*, 15:64-79 (July, 1924).

[10] Of course, this is put in extremely simple terms. A method would be better which did not get quite as many things learned but which secured this somewhat inferior result at *very* much less effort, or a method would be better which got considerably more things learned with not much greater effort.

Measuring learning: retention. But these two methods of measuring learning have reference only to immediate memory. Although we have large and important uses for immediate memory, much of the learning we do in life and most of that which is done in school is for more or less permanent retention. We say we want children not only to learn things but to remember them. Accordingly, another way of measuring the effectiveness of memory work is to find out the amount of the material that can be reproduced after a stated interval. All sorts of intervals have been used in experimental work from ten minutes to two years.

A difficulty here, however, is to measure the amount retained over the period in question. In the case of verbal material some have done this by counting the number of correct words reproduced. There are obvious disadvantages in this. Some words are more important than others. Omissions are different from substitutions, yet they are counted the same. Extra words may be introduced, but the method takes no account of them. The words may be correct but in the wrong order without affecting the score. Others have sought to measure retention by dividing the passage into ideas rather than words and counting the number of correctly reproduced ideas. Aside from the fact that no two people would be likely to agree as to the division into ideas, this method is better—though not much better—than the word-count method. Thus the measure of retention according to the amount of material reproduced after an interval, though simply expressed and easily understood, turns out to be difficult to apply.

The saving method. There is another important way of measuring learning from the point of view of retention. This method takes account of the time or the number of repetitions required to *relearn* the material after it has been dropped and at least partly forgotten. This is known as the "saving method." The less time it takes to relearn a stanza of poetry two weeks after it was first learned, the better the retention is.

You will find that this saving method is an interesting way of securing for each pupil a record of his learning ability. The application of the method, however, depends upon having at hand learning tasks of equal difficulty. Suppose you were to set up as an objective the securing, both for present and future use, of learning tasks of equal difficulty. With reference to poetry one way to proceed would be to select a long poem or several short ones such that the material could be broken up into stanzas or sections of the same length and *probable* difficulty to learn according to your judgment. Then try out a number of stanzas with your pupils. They may work in pairs, one pupil "hearing" the other as soon as he thinks he can repeat the stanza. With little children it may be necessary for the teacher to be the hearer. Reject the stanzas which turn out to be unusually easy or difficult and save for future use a dozen or so which it has taken the children about equally long to learn.

With this material in hand you are prepared to test the learning ability, for this kind of material, of another group of children. You might begin by having the children learn one of your "standardized" stanzas, taking for each child the time required by him to attain an errorless reproduction. Do this for a second stanza and continue to use different stanzas until you feel that by averaging the times required by each pupil you have secured a reliable measure of his rate of learning one of these stanzas. This ought not to use up more than five or six of your selected stanzas. Now take an additional stanza and give all the pupils the same length of time in which to study it. This length of time should not be greater than the average time required by the most rapid learners in learning the stanzas on which you have already obtained a record. You might, for example, allow all the pupils three minutes. At the end of this time the pupils should turn the papers over, and the copies of the stanza should be collected. At a convenient time later in the day or the next day the same stanza may be given to the pupils to learn to the point of errorless reproduction. If it takes a child four minutes to do this, whereas his average rate of learning stanzas of equal difficulty as already established is six minutes, the saving for him due to the first (three-minute) learning period is two minutes. This may be taken as the measure of his learning in the three-minute period.

Or you may have your pupils learn one of your selected stanzas to the point of errorless reproduction. Then after an interval, say of a day or a week, you may have them relearn the same stanza to the same point. The amount of time saved is the measure of retention of the first learning. For more reliable results this procedure may be repeated several times, and an average time saved may be computed for each pupil.

The saving method may, of course, be applied in measuring the learning of other kinds of material. You may, for example, provide yourself with a selected vocabulary of French words of equal difficulty to learn, that is, to learn with their English equivalents. (You will no doubt have to try out a longer vocabulary with *other* children in order to secure this selected vocabulary.) You may then have a set of 20 words learned and relearned after an interval. The difference between the time of first learning and the time of relearning is a measure of the goodness of the original learning from the point of view of retention.

The part and whole methods in immediate memory. When I said above that the evidence pointed to the superiority of the whole method over the part method of learning, I evidently made a general and certainly not a very critical statement. Let us look at some of the details as they relate to *immediate memory*. Pyle and Snyder find that for passages of poetry of varying lengths "whether five lines or 240 lines . . . are memorized, learning by whole is without any exception more economical than learning by part."[11] This verdict is pretty generally agreed to. I shall not burden these pages with a large number of citations, but perhaps one from an entirely different source may be of interest. Des Bancels says: "It (the whole method) requires less time and fewer repetitions than the part method. The difference is not very great, but it is entirely consistent and is true of every individual. Moreover it exists whether the material is poetry or nonsense syllables."[12] Winch's evidence, as I have already pointed out, is to the contrary; but the consensus is altogether too great to be set aside or even brought into serious doubt by a single investigator.

The general superiority of the whole method in respect to immediate memory may, however, be seriously affected and indeed in some instances lost. Four conditions are worth considering: (1) the kind of material learned, (2) the length of the material, (3) the uniform difficulty of it, and (4) the individuality of the learner. These modify in specific instances the application of the general rule. In the first place, as to the kind of material, the clearest evidence in favor of the whole method has been secured when the learning task was of a rather formal character. Syllables, digits, vocabularies, and poetry and prose selections constitute the usual materials. All these present tasks of a mechanical character. It is not so much that the method does not apply to other kinds of material as that it has not been applied. The experimentalists have not utilized to any great extent the other types of learning which both in school work and in life are so frequently employed. As to the application, therefore, of our rule to different kinds of material, we can only say that it applies *so far as it has been applied.*

---

[11] Pyle, W. H., and Snyder, J. C. "The Most Economical Unit for Committing to Memory," *Journal of Educational Psychology*, 2:133-42 (March, 1911).

[12] Des Bancels, J. Larguier. "Sur les Méthodes de Mémorisation," *L'Année Psychologique*, 8:185-204 (1908).

In the second place, as to length of material, the results are to the effect that the longer it is the more pronounced the benefit of a whole method is. For example, Pyle and Snyder in the article already quoted say: "The relative saving is much greater in the case of long selections that require more than a single sitting." This statement may be taken as typical of the conclusions reached by other investigators. It is to be understood, then, that as far as immediate memory is concerned the whole method yields better results when the material to be learned is long. At this juncture, however, common sense must come to our rescue. It is entirely appropriate for us, like Watt,[13] to give our adherence to the whole method, while at the same time refusing to make it "go on all fours." No one would advocate the learning of *Paradise Lost* by reading it through from beginning to end until it was learned. In very long pieces the whole method will have to be modified by dividing the material into tolerable lengths. We must, as Watt goes on to say, come down to the level of human power and encourage the will as much as possible. For as everybody knows and as has been abundantly shown by the experimenters, the learner "just naturally" adopts a part method if he is not otherwise encouraged or directed. The principal reason for this seems to be that he can see his progress when he learns a little at a time.

---

[13] Watt, Henry J. *The Economy and Training of Memory.* New York: Longmans, Green and Co., 1910. pp. 48-55.

In the third place, as to uniformity of difficulty in the learning task, the evidence is that the whole method justifies itself most clearly when uniformity is greatest. If there are spots of particular difficulty, either for persons in general or for the particular learner, it appears that a strict application of the whole method is not advantageous. Special attention to the difficult parts is desirable.

Finally, the whole method appears to give the best account of itself among the more capable learners, although less capable learners may acquire through practice a facility in the use of the whole method which may make it superior even for them.

Evidence as to retention. But what about the whole and part methods in relation to permanent retention? Most of the things we learn we wish to retain. Does the whole method work out better than the part method from that point of view? It does; and the evidence appears to be even stronger than it is in the case of immediate memory. Indeed, some individuals can recall better immediately after using the *part* method of learning—especially if the material is short—than they can after using the whole method; but such exceptions to the superiority of the whole method have been very seldom reported when the learners have been asked to recall the material after a long interval. If therefore the learning task is short and the recall immediate, the part method may sometimes be as good or better; but as the material lengthens and as the period of recall likewise lengthens, the superiority of the whole method becomes more pronounced. It appears also that the whole method gains more from practice than does the part method. Even if pupils do better at first with part learning, they quickly improve in whole learning and surpass their efforts at part learning. The experiments of Lakeman are typical in regard to these matters.[14]

The advantages of a mixed method. But to say that the whole method is better than the part method is after all only to make a rather general statement. The whole method *in its extreme form* is not often the best method. It assumes among other things that the material is of fairly uniform difficulty throughout. If such is not the case the difficult spots should be located and given separate attention until they are raised to a point where to complete the learning of them is a task of about the same difficulty as the learning of the rest of the material. In other words, since the whole method proceeds more or less on the assumption of equality of difficulty, the learner should make the material present this equality by giving early and special attention to difficult parts. Moreover, it is admitted that, with all its disadvantages, the part method *does* take advantage of the nature of attention. We know that concentrated attention is impossible except for short intervals. We know that a task attended to for a relatively long time "sags in the middle." This is made evident by the commonly observed fact that the middle of a long task is not as well learned as the two ends of it. The part method, by giving attention to short units of work, tends to obviate this difficulty. The suggestion is, therefore, made that this advantage be secured for the whole method by dividing the work into parts, marking them off quite definitely one from the other, and in the learning process pausing between the parts thus set up long enough for the revival of attention (Meumann's "mediating procedure").

---

[14] Lakeman, Mary E. "The Whole and Part Methods of Memorizing Poetry and Prose," *Journal of Educational Psychology*, 4:189-98 (April, 1913).

Controversialists may say that we have here neither a part method nor a whole method, and that to assert the superiority of a whole method when thus modified is to juggle with words. But we are not controversialists, and we do not care in the least whether the whole method is better than the part method or whether the truth is the other way around. What we are interested in is the question of the *best* method in a given situation. If that best method is an intermediate one—and it unquestionably is—its elements may classify it with the whole or with the part procedure. At present it looks very definitely as if a mixed method is better than either pure method, but that the mixed method which will most frequently give satisfaction contains more elements in common with the whole method than with the part method. Moreover, it is almost certain that if we train our pupils (or ourselves, for that matter) in whole learning we shall find it more and more satisfactory in its purer forms as training continues.

I think, then, we may fairly conclude that for the purposes of the school the whole method of learning—meaning by that the various methods dominated by the whole idea—is better than the part method. If a pupil has to prepare a textbook assignment he will undoubtedly do better if he reads it over carefully and re-reads it, each time in its entirety, than he will if he breaks it into parts, re-reads the parts, and then puts them together. He will probably do better still if after his first careful reading of the entire assignment he picks out, not parts which when strung together make the whole, but parts which present difficulty or which are particularly important, devoting attention to these parts before continuing the reading from beginning to end.

Moreover, there are larger implications in the evidence. What applications we may ask ourselves, does the "whole" idea in learning have to those larger masses of material which constitute the task of a year or of a semester? In other words, what bearing if any has this doctrine on the way the course of study shall be taught? It *suggests* that, in teaching the subject-matter of a course, more advantageous learning will be secured if the teacher first takes the pupils over the entire course rapidly, selects difficult and important parts for special treatment, and then covers the entire material a second or third time, adding further details, new points of view, additional applications, and wider generalizations.

The need for more practical investigation. But in all this we cannot really be sure. The psychologists have spent an inordinate amount of time in their learning experiments dealing with nonsense syllables and with the learning of poetry and prose passages. I should be the last to deprecate this. There is, for example, a real reason for the experimental investigation of memory on material which is devoid of meaning—the dead lift of memory, as it were.

But all the reasons for this procedure have reference to it as a point of departure; and to remain on the level of meaningless material is to be forever beginning and forever failing to attack the real problem. Logical memory, association of ideas, and the apprehension of meanings—these are specially emphasized in the modern school. I, therefore, agree with Kuhlmann, and with Pyle and Snyder who quote him, that more practical results would be obtained if the investigations of memory involved learning and memory processes as they are found in the schools and in life. Who is in a better position to make such investigations than the classroom teacher?

FORGETTING

After a thing has been learned, we know that forgetting sets in. The psychologists have much to say about forgetting, or, as some call it, obliviscence. Its opposite is retention, and I have already intimated that those who learn rapidly tend to retain well. Some other statements that may be made on the basis of evidence are:

(1) People differ less in their rate of forgetting than in their rate of learning. In other words, when material is once learned it stays in mind to more nearly the same extent for everybody. *Therefore, the original learning is the important thing.* You should know your pupils and knowing them insist that each of them try hard enough and intelligently enough to fix his work at the first learning.

(2) Every one of your pupils probably has his own *learning index* which expresses the time he needs to learn a given type of material to the point of reproduction. No doubt

he also has his index of forgetting. It is as desirable—perhaps more so—that you should know his indices of learning and forgetting as it is that you should know his score in a standardized arithmetic test.

(3) Memory, both immediate and deferred, improves by practice in the field where it is exercised. Psychologists are pretty well agreed in their conclusions that memory as a general capacity cannot be improved, but memory for particular materials can be markedly improved.

(4) Forgetting can be offset by relearning which, however, may take place at longer and longer intervals. The following statement from Lyon is offered as typical. A person after memorizing a poem of thirty stanzas obtained the highest retentiveness by distributing subsequent readings according to the following intervals: 2 hours, 8 hours, 1 day, 2 days, 4 days, 8 days, 32 days, etc. This suggests, and Lyon supports the suggestion, that the intervals obey a rather definite law of geometrical progression.[15]

How to counteract forgetting. There is practical meaning in this—and some problems. The meaning is that after pupils have learned a thing they should be trained to recall it shortly after the original learning. In order to make it a permanent possession—in order to offset forgetting—additional recalls should be made. But the intervals between these recalls may be longer and longer as time elapses. The rule is to review the material whenever it starts to fade; and it appears that these occasions will arise at intervals which exhibit a rather definite progression.

---

[15] Lyon, D. O. "Relation of Length of Material to Time Taken for Learning and the Optimum Distribution of Time," *Journal of Educational Psychology*, 5:161 (March, 1914).

Probably the nature of this progression is different for different pupils; and this introduces us to one of the problems. If Lyon could ascertain by repeated trials the best distribution of recalls for an individual, a teacher could do the same for at least some of the pupils—the problem cases—in his class. It may be assumed, first, that upon investigation a teacher will find that there are certain broad general truths regarding the best intervals at which recall should take place; second, that there will also be specific truths of like import but different in detail for individual pupils. An important problem connected with teaching will then be the distribution of drill or review so as most nearly to correspond to the intervals thus established. This is a real problem, and it involves in particular such questions as this: After a child has learned the simple addition combinations, how much drill should thereafter be provided and how should the recurrences of these combinations be distributed? When teachers are able to answer, even in general terms, problems like this, expertness in regard to teaching will reach new and higher levels.

Meanwhile, pending more accurate determination, a general rule may be suggested with some assurance. In your own study as well as in that of older pupils who have to handle long textbook assignments or take notes on lectures, it is good procedure to review the material shortly after the first learning took place. It is productive of decidedly beneficial results to go over one's lecture notes immediately after making them and then if possible to review them a second time later on the same day. After this is done, forgetting will be so far forestalled that the lapse of a week or two will not then make much difference.

Reminiscence. With further reference to this matter of forgetting one may at least call attention to the opposite tendency, which has been technically called reminiscence. With some learners and with some materials—notably those which have for their object the establishment of skill—it has been frequently shown that with the lapse of time the learner's stock of knowledge or skill is actually increased. For example, in a series of drill periods certain persons will begin a given period with a higher rate and a greater accuracy than they have exhibited in any previous period. Investigations of the effect of the summer vacation on the knowledge of school subjects has brought out the fact that some children give better test results in September than they did in June.[16]

We need to know more about this phenomenon of reminiscence—whether it depends upon any particular technique in the original learning, or upon the completeness of the original learning, or upon the renewal of the material in periods subsequent to the original learning and prior to the one during which reminiscence is observed; and whether the person in whom this phenomenon occurs has in reality "thought about" the material during the interim. We should also like to know what types of learning material are favorable to this phenomenon. In short, there is a field for investigation, both for psychologists and for teachers, as to the conditions under which the phenomenon of reminiscence occurs and the means by which its benefits may be secured.

---

[16] Morrison, J. Cayce. "What Effect Has the Summer Vacation upon Children's Learning." *Educational Research Bulletin,* 3:245-49 (October 1. 1924).

TRANSFER OF TRAINING

It is of profound pedagogical importance to know the extent to which learning of one kind influences learning of another kind. One aspect of the matter has been extensively investigated under the heading of "the transfer of training." The form in which this problem has been set up is narrower than that which I propose to give it in this connection. One of the historical experiments will illustrate what I mean. Winch trained a group of children in the solution of reasoning problems in arithmetic, and then sought to measure any increase in their ability to reason about non-arithmetical matters. He obtained a small increase in the ability in which the children were not trained.

Transfer as facilitating learning. But this is not directly a learning problem. It would become so more directly—and perhaps less directly a problem in the transfer of training—if Winch had sought to find out whether the children, after having been trained in the first activity, could be *taught* the second activity more easily or more quickly. There may be, in a real pedagogical sense, transfer from one situation which has been learned to another which has not been learned, without the transfer manifesting itself in an immediate response. An analogy will show what I mean. Suppose a poem to have been learned years ago, and suppose that the learner is now called upon to reproduce it. He may find himself quite unable to do so. We do not say, however, that there is no present effect of the original learning, because he can undoubtedly relearn the poem in much less time than it took him to learn it originally. In strict literalness nothing that is ever experienced is wholly forgotten. Many an adult has no visible knowledge of the Latin or the physics which he "took" in his student days, but he could relearn these things with far greater ease than he could learn them if he had never studied them.

So, it seems to me, it is with the transfer of training. The question whether after learning one thing a child can do another is different from the question whether after learning one thing he can learn another more easily. The best evidence is that to deny the transfer of training, even in the narrower sense, is wrong. All competent authorities are now agreed that some transfer exists, and the pedagogical question is, How may we take advantage of this phenomenon?

The circumstances under which transfer takes place. It is apparent that the more closely the thing which has been learned resembles the thing which has not been learned the more the ability to learn the latter will be favorably influenced—in other words, the greater the transfer effect. It is apparent, too, that this resemblance which permits transfer to take place may not only be in the materials themselves, but also in the methods by which they are learned, the ideals, attitudes, and emotions by which they are dominated, and in short by every element, both in the learner and in the things to be learned, which exists while the learning is going on. Moreover, if these similarities are brought to the attention of the learner, they are much more effective in producing improvement in the new situation than they are if he is unconscious of them. For example, when Bagley taught children to be neat in writing their arithmetic papers, and observed no improvement in the writing of language and spelling papers, the trouble was that neatness as a general ideal had not been raised to the consciousness of the children. Neatness to them was neatness in arithmetic, and as such it was desirable; but neatness in language or spelling was another story. When, however, Ruediger repeated Bagley's experiment, he found that emphasis on neatness in a single subject when accompanied by talks to the children about neatness in general produced a decided improvement in the appearance of papers in geography, arithmetic, grammar, and history.

On the other hand, it is well to emphasize the fact that much of our teaching is carried on as if transfer to unlearned activities took place to a far greater degree than is the case. We cannot train the memory so that learning of *any kind* of material may be facilitated. There is no "faculty" of discrimination such that training in detecting small differences in musical pitch has any useful effect on the ability to distinguish colors, or the ability to judge distances or durations of time.

In our arithmetic teaching we have abstract work and problem work. The latter involves reasoning, the former involves computing ability. Both are in the subject of arithmetic as we teach it. They may, therefore, seem to be so closely related that transfer of training from one to the other would be extensive. Yet in repeated experiments under the most carefully controlled conditions, Winch found that drilling school children in computing, though it greatly increased their ability in that respect, had no clearly observable effect on reasoning in problem work. Thus it appears that "ability in arithmetic" is not *one* thing. Indeed, if the group of abilities and skills to which we give the name "arithmetic" is analyzed, it will be found that there are hundreds if not thousands of more or less specific abilities covered by this term. Moreover, matters outside of arithmetic may be more closely connected with matters included in this subject than the latter are with each other. Winch also found that when he trained children on reasoning problems in arithmetic to the point where they showed an improvement of 150 per cent he actually improved their reasoning on non-arithmetical material about 30 per cent. In other words there is enough similarity between arithmetical and non-arithmetical reasoning for an appreciable amount of transfer to take place. We must not be misled by names. They are often used to cover widely different—that is, psychologically different—types of knowledge and skill.

Practical meanings. The pedagogical implications of the fact that transfer is partial, but not total, are clear in their general outlines. If we wish children to improve in a particular way, our surest means of causing them to do so is to train them in the specific ability we have in mind. If we desire to improve their rate of reading to the greatest degree, we should attack reading rate and not something else. On the other hand, the evidence is that such a procedure will influence favorably their comprehension of reading, although not nearly to the same extent. In arithmetic we have what we call the 45 basic combinations in addition.

Teachers often act as if they supposed that the learning of these combinations would "spread" to all the necessary combinations in addition, that is, to all those required in column addition and in carrying in multiplication. This is not the case.

In the first place, research has shown that not 45 but 100 combinations must be regarded as basic. For example, 2 + 6 and 6 + 2 are not the same. Clapp[17] has shown that among 1154 third-grade pupils 220 got 2 + 6 wrong, whereas only 139 of the same children got 6 + 2 wrong. Evidently the learning of 2 + 6 is no assurance that 6 + 2 is likewise learned. Since when we speak of 45 basic combinations we assume that these two facts and all other pairs of reversible facts are the same, and since the evidence is to the contrary, it is better to recognize all these as separate facts; and this will yield, when the zero combinations are included, 100 basic number facts. I do not mean that there is not a pronounced transfer from 2 + 6 to 6 + 2, but I do mean that after one has been learned the other will have to be separately learned, although very likely with less effort, especially if the two facts are associated in the teaching and learning.

In the second place, the basic combinations, even when expanded to one hundred in number, are not sufficient for general use in addition. A pupil may know 2 + 6 to the point of an invariably correct response without knowing the derived facts, 12 + 6; 22 + 6; 32 + 6, etc. Osburn[18] points out unmistakably that each of these derived combinations is a new fact to be learned. No doubt a knowledge of 2 + 6 will facilitate—that is, will transfer to—the learning of the corresponding derived facts, but it is doubtful if any but the most unusual child who knew no more than the basic combinations would without *additional* learning be able to infer the derived facts.

---

[17] Clapp, Frank L. *The Number Combinations, Their Relative Difficulty, and the Frequency of Their Appearance in Textbooks.* University of Wisconsin, Bureau of Educational Research, pp. 22-23 (Bulletin No. 2, July, 1924).

Accordingly, it is clear that you as teachers must in all your teaching steer between the assumption of total transfer and the assumption of no transfer. You must—and I think the evidence bears me out in this—recognize that in teaching there are two factors to be considered, namely, a general factor and a specific factor. In the illustration from arithmetic, the general factor will tell us that in respect to the derived addition combinations the learning has been partly achieved through the previous learning of the basic combinations. At the same time the specific factor will tell us that this is not sufficient, that there must be a separate learning process with reference to the derived combinations themselves. The general factor corresponds to the transfer of training; the specific factor to the necessary direct frontal attack on each particular item to be learned.

Your task should be, in the first place, to teach basic facts and principles in such a way that they will be of the greatest possible potency in assisting children to learn new facts or to meet different conditions. Methods of work will, therefore, be important in your teaching. You will not expect to train the memory, but rather to inculcate economical methods of study. Again, you will seize upon the general features of a learning process—aside from those of method—and make children aware of them. For example, you will not teach neatness as something that belongs to arithmetic; but on the other hand, although you may be a teacher of arithmetic and not of other subjects, you will not fail to bring the general ideal of neatness in all school work into consciousness. Whether this will, unaided, have an appreciable effect on the appearance of the geography or history papers may or may not be true. That question is largely academic, but the evidence warrants us in believing that the geography or history teacher will be able to get results more easily, because you have set up this general ideal in the arithmetic class.

---

[18] Osburn, W. J. *Corrective Arithmetic.* Boston: Houghton Mifflin Co., 1924. 182 pp.

In the second place, you will not fail to teach directly, even though the teaching may be very brief, the particular knowledges and skills which the school requires. You will be prepared to expect that neatness when taught in connection with arithmetic will not cause the grammar papers to be sufficiently neat. The conditions of neatness in grammar are different from those in arithmetic. The ideal is the same; but its application is a new fact, or a new series of facts. Satisfactory results will be secured only when these facts are taught.

The determination of the extent to which in given situations transfer from old learning may be relied upon and of the extent to which additional teaching is demanded, requires the exercise of fine professional judgment. At present the results of research will help only a little. You should go about your teaching with open eyes, and with a mind alert to inquiry of this sort. You should carefully note and record the extent to which in each topic you have been able to rely upon what I have called the general factor in learning and the extent to which specific learning was required.

This spirit of inquiry, the spirit of open-minded alertness to the problems which arise in teaching, is the attitude for which I am pleading. This attitude will lead you to seek problems, and to do something, even though it may be but a little, in their solution. It will, I hope, cause you to organize your class in a definite attack upon a few of the problems and will cause you to record the results of your efforts. It should be the aim of all good teachers not only to become experts in the training of young people, but also to leave some record of their experience for the guidance of others.

The psychologists have extensively studied the question of transfer of training, and they have carried us to the point where we may feel confident that transfer takes place. They have likewise given us analyses of the conditions under which transfer may be expected. But their work stops at the classroom door. It remains for the teacher to examine the question from the teaching point of view. In particular, it is for the teacher, with his children as subjects, to ascertain with respect to definite situations how much transfer can be brought about by an appropriate method of teaching, how far it is reliable in a new situation, and how far a direct teaching attack upon the new situation is required for coping with it successfully.

TYPES OF LEARNING

Not only are there pronounced individual differences in rate of learning and in accuracy of reproduction, but there are also fundamental qualitative differences in learning. The objects about us appeal to us in various ways. We are endowed with sense organs through which we become aware of these objects. Some objects appeal to us through the eye, some through the ear, and still others through other senses. Sometimes a given object presents itself simultaneously to more than one sense. When we meet a person, we ordinarily have sensations of form and color through the eye and of the sound of his voice through the ear.

One of the most important senses is the muscular sense. We say we get the "feel" of things. Most of us can write our names as well in the dark as we can in the light. While the muscular sense is not localized in any part of the body, the hands and vocal organs are particularly important. Thus in the muscular sense we distinguish the manual-muscular and the vocal-muscular.

I have said that there are qualitative differences in the way people learn. The so-called "ideational type" of the individual is one of the bases of these qualitative differences. Some people see better than others, not that their vision is more acute, but that they make better use of the visual images which they get from seeing. But more important than a comparison of the strength of visual imagery among different individuals is the comparison of visual with other types of imagery in the same person. For what has just been said as to the meaning of visual imagery may be said with respect to auditory and motor imagery. Now, if a person has better visual imagery than any other, he tends to use this imagery in his thinking and in his learning, and to use it whenever he can to the exclusion, or at least to the subordination, of the other types. The same is true with respect to auditory imagery. The person in whom this type of imagery is strongest thinks and learns through the use of this type of imagery. He remembers people by their voices. Rhythm and timbre are excellently retained and utilized in learning. Again, the person who has strong motor imagery tends to gesticulate, and especially to feel and to half-formulate words in the vocal organs. We note this especially in the lip-movement which characterizes the learning of young children.

A method of investigating types of learning. Years ago much more was made of types of imagery than is now the case. In those days the doctrine was wrapped up with phrenology and has since been partially discredited along with it. Not a few writers, however, within the last ten or twelve years have pointed out that this question of types has important practical meaning. Methods have been devised for determining types, and although these methods are far from infallible they are nevertheless useful. One of the most important of these is the *method of helps and distractions.* As regards helps, this method proceeds on the assumption that the visual learner will be assisted by visual presentation and by supporting devices of a visual character; that the auditory learner will be similarly helped by auditory presentation and the motor learner by presentations involving muscular movements. As to distractions, it is held that the visual learner will be hindered in his learning by the introduction of difficulties and extraneous appeals of a visual character, whereas the same learner will be relatively less affected by distractions of an auditory or motor character. Similar statements would hold as to the distractions which would affect the auditory and motor learners. Each, it is held, will be pronouncedly affected in his learning by interferences of the kind belonging to his preferred type of imagery.

Pure learning types seldom, if ever, exist. The person who can learn by but one of these methods is rarely found. On the other hand, most people have a preferred or dominant type of learning; and the method of helps and distractions will assist in determining its existence. If, for example, we wish to establish by the method of helps the presence or absence in a given child of a preferred method of learning lines of poetry, we may present the material to him in a number of different ways. (Again we shall have need for selected stanzas or sections of equal difficulty.) We may require the pupil to read the material silently, thus appealing directly to a visual mode of learning; or we may read the poem aloud to him, thus making an auditory appeal. We can scarcely present the material by a purely motor method, but we can use this method in combination with the auditory or the visual. If, for example, we allow the child to read silently and to whisper the words as he does so, and if we find that this materially helps his learning, it is evidence that he makes effective use of vocal-muscular response. Again, if writing the poem is of material assistance, there is some evidence as to his use of manual-muscular methods in learning.

If the child learns better by the visual appeal than he does by the auditory appeal, the evidence is that as between these two modes of learning he prefers the former. In the auditory appeal he has either been obliged to work by a method which he does not naturally use or else he has been obliged to translate the heard words into seen words. In either event his learning is presumably easier when the poem is presented visually.

We may now interpose one or more distractions. Suppose we have the pupil hold his tongue between his teeth while learning one of the stanzas. This will suppress internal speech and will make difficult not only the operation of the motor learning connected with the speech organs but also to some extent the operation of auditory learning (inner hearing). If the pupil's memorizing under these circumstances is not much impaired, we may infer that he learns well by the visual method—in other words, that he is not much dependent upon the assistance of muscular and auditory learning. If his memorizing is markedly impaired by this suppression of inner speech, he is probably one of those interesting cases which so often constitute school problems. If noises are especially distracting to him in his learning, it is believed to be an indication that auditory learning is an important mode of procedure with him. If the sight of other objects than those connected with the learning interferes, the evidence is in favor of a visual type.

If visual learning is facilitated by a definite arrangement of material in the form of groups or patterns, the indication is that the learner favors a visual procedure. Other indications as to types of learning may be noted. The strongly visual learner can reproduce nonsense material, for example letters, in an inverted arrangement practically as well as in a direct arrangement. The person who is dominantly auditory- or motor-minded cannot do this. In learning words the visualizer confuses those that look alike while the auditory learner confuses those that sound alike.

I suggest that at least in the case of pupils who are not maintaining themselves in your classes you take some steps to find out whether they have a preferred mode of learning and if so what it is. You may at least present your teaching material by different methods and see whether they learn better in one case than in another. In support of this idea I call your attention to the report by Fernald and Keller which appeared a few years ago.[19] Certain children of normal or superior intelligence had been quite unable to learn to read in school. They were already nine or ten years old. One of them was over twelve. They could do well in arithmetic, history, or geography if they could have their lessons read to them, but read they could not and never had. Their teacher believed them to be stupid. Investigation proved, however, that they were well endowed mentally. The report by Fernald and Keller tells how these children were taught to read.

Now, our ordinary methods of teaching reading are dominantly visual, and this is undoubtedly as it should be. There is likewise, especially with beginners, a strong emphasis on auditory methods; and no doubt this too is correct. But since these children had failed to learn to read after three or four years of effort and since they were of normal mentality or better, Dr. Fernald inferred that their mode of learning was dominantly of the motor, or kinæsthetic, type. Accordingly, she prescribed a method of motor learning. It was then found that these children could readily learn new words if they wrote them—thus employing a manual-motor method—and at the same time said the syllables to themselves—thus employing a vocal-motor method. Moreover they could afterwards recognize these same words in print. There was more to the method than this, but the appeal throughout was to the motor elements in learning. The children rapidly learned to read by this special method—a method addressed to their preferred mode of learning. It was noted, however, that these children, even after they had learned to read well, tended to write out and to spell and pronounce aloud the new words which they encountered.

---

[19] Fernald, Grace M., and Keller, Helen. "The Effect of Kinæsthetic Factors in the Development of Word Recognition in the Case of Non-Readers," *Journal of Educational Research*, 4:355-77 (December, 1921).

You will note that it took three or four years of failure in reading and, at the end of that period, a rather peculiar combination of circumstances involving, among other matters, the presence of a trained psychologist to bring out the fact that these children were motor-minded. The teachers who declared these pupils to be mentally deficient might have detected their preferred mental behavior if they had been better informed about these matters and had used some such methods of helps and distractions as I have described.

I have given this illustration because I think it brings out an important fact—namely, that many children are being forced to learn by methods which are difficult for them to employ. I believe we have many children in our schools whose intelligence is good but whose mode of learning has not been considered in the mass instruction to which we are committed. As far as group teaching is concerned our safest method is the multiple appeal, the presentation wherever possible of the material in visual, auditory, motor, and tactile ways. But this is not sufficient. Some individual teaching will have to be done, both on the grounds set forth in this connection and for other reasons.

You may feel that you have not time to do this. Yet it is certain that some teachers, even with large classes, are doing this very thing. See, for example, the report of Ruth Geiger, a third-grade teacher in Cincinnati, on individualizing the work in reading in a class which at no time of the year numbered less than forty-five.[20] Moreover it is equally certain that many teachers, even when they have small classes, are failing to make the most of their opportunities.[21] In most cases this is, in my judgment, because these teachers do not know how to proceed. In particular they do not know how to locate the deficiencies of individual pupils and to apply the necessary remedies. A part of the purpose of this book is to bring to the attention of teachers some of the work which has already been done along this line. The chapters on "Error and Failure" and "Reaching the Individual" are largely devoted to this matter of individualizing instruction.

[20] Geiger, Ruth. "A Study in Reading Diagnosis," *Journal of Educational Research,* 8:283-300 (November, 1923).

[21] This is one of the permissible inferences from the six or eight important studies of the effect of size of class on the efficiency of teaching. The evidence is to the general effect that the teachers of small classes are securing about the same results as teachers of large classes.

MENTAL FATIGUE

Possibly there is no aspect of mental work about which we know less than we do about fatigue. This seems strange, since the phenomenon of fatigue appears to be generally experienced and to present a familiar aspect. Part of the trouble is that we do not critically state the meaning of the term. Fatigue may be properly defined as a diminution in productivity *which rest alone can restore.* We do not ordinarily distinguish between fatigue as thus defined and certain feelings of ennui or distaste which accompany prolonged mental work. That these feelings are by no means to be identified with diminished powers of production is made clear by the fact, first, that the feelings may themselves be postponed or counteracted by interest and incentives, and second, that production often continues unabated in spite of these feelings. If productivity is restored by incentives, it is clear that the diminution was not directly caused by the prolongation of the work, and that the productivity was certainly not restored by rest. It was reduced by the weakening of the stimulus to productivity and restored by the strengthening of that stimulus. Again, even after mental work has been long continued and when pronounced feelings of fatigue are experienced, the actual product may remain unaffected either as to amount or accuracy. In experimental work it has frequently been found that the person who was working thought he was tired out but that he was actually continuing the task with undiminished efficiency.

Fatigue, therefore, which represents a reduction in productive power due to a continuance of activity is one thing, and fatigue as a reduction of product due to feelings of repugnance to continue the work is another thing. You may say that the result is the same in both cases, namely, a diminished product. Such is indeed the case. But the only cure for real fatigue is rest, while the cure for a "feeling of fatigue" is the removal of the feeling and rest is only one way in which this may be done. The distinction is very real and has been repeatedly demonstrated experimentally. Feelings of fatigue are banished by introducing additional rewards or incentives.

There is, then, a real distinction between actual fatigue and those emotional states which we frequently identify with it. This distinction is most manifest when we consider the treatment to be prescribed. For real fatigue rest is the only cure. Feelings of fatigue may be due to the superior attractiveness of *something else* than the work in hand. The child wants to quit this work, because he wants to do something else. He will rightly tell us that he is "tired of" the work rather than "tired by" it. In reality he has been "tired" as much by what he is not doing and wants to do as he is by what he is actually doing and does not want to do. His so-called fatigue is a distaste for the matter in hand and this may easily arise from the superior attractiveness of some other activity.

Mental fatigue may generally be disregarded. The results of experiments justify me in advising teachers that for the most part they may disregard mental fatigue—that is, fatigue in the real sense as opposed to feelings of fatigue. Maintaining the proper mental hygiene of the classroom means eliminating emotionalisms and worry far more than it means avoiding fatigue. It is much more dependent upon the skillful use of worthy purposes, knowledge of success, and right incentives than it is upon the reduction of otherwise desirable periods of work or the introduction of not otherwise desirable periods of recreation. I say "not otherwise desirable," for recreation and play may rest upon their intrinsic value and justify the time devoted to them without reference to their therapeutic properties as cures for fatigue.

One large item in connection with fatigue is the fact, brought out by Heck in this country and by Winch in England, that in the ordinary school day one time is about as good as another for mental work. We commonly suppose that children are at their best during the earliest hours, and we, therefore, tend to place the hardest subjects at the beginning of the daily program. Winch showed, among other things, that work done in arithmetic during the hour from eleven to twelve o'clock in the morning was better than work of the same kind done during the hour from nine to ten. Heck, after experimenting at different times with children in New York, Lynchburg (Va.), and Roanoke (Va.), formulated his conclusion as follows: "Normal, healthy children of the grammar grades, in a hygienic school environment, can meet the requirements of the usual daily school program without injury to themselves or their work." In another connection the same investigator puts his findings in a little different form when he says, "Mental fatigue is negligible in school management, provided the children are working under proper hygienic conditions."

It appears, therefore, that among the "things we know about learning" are these: that mental fatigue is not the bugaboo in school that many have supposed it to be, that teachers may well banish boredom by making the matter in hand more attractive, and that children—except possibly the youngest—can negotiate the school program about as well at one time of the day as another. This, of course, is on the assumption that they are in normal physical condition and that the environment is, as Heck says, hygienic. I cannot too often nor too earnestly urge you to try out these things for yourself. There will always remain a "reasonable doubt" concerning many of the truths we know or think we know about learning until teachers carry the demonstrations of these truths into the classroom. Try the experiment of restoring mental activity through new incentives. If you have a feeling that arithmetic ought to come at nine o'clock or at any other particular time of the day, try it at other times and see whether you are right about it or whether the conclusions of Winch and Heck apply in your class.

THE LEARNING CURVE

My final point in considering some things we know about learning has reference to the nature of the so-called learning curve. This idea of a curve is, of course, derived from the mathematical treatment of learning. Let us suppose that a diagram is to be made of a certain child's improvement in multiplying three-figure numbers by two-figure numbers as the result of five minutes of daily practice. Along the base of the diagram we may arrange spaces for the days—at the left the first day, then the second, third, and so forth. On a vertical at the left we may lay off as measures of success the number of correct answers secured at each practice period. The smallest measures of success should be at the bottom and the largest at the top. The diagram before the record of the child's work has been put in may look like Figure 1.

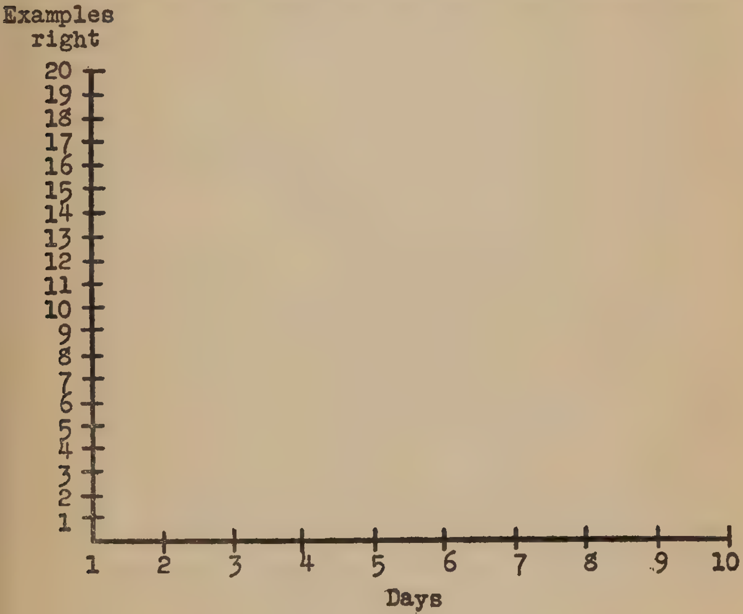

FIGURE 1.

Suppose the record of a certain pupil for the first ten days was:

DAYS	1	2	3	4	5	6	7	8	9	10
EXAMPLES RIGHT.	7	10	10	12	10	14	14	15	13	15

Then the curve of learning may be inserted in Figure 1 with the result as shown in Figure 2. Vertically above the point for the first day a circle is placed at 7 to indicate that on that day the pupil had 7 examples right. Vertically above the point for the second day a circle is placed opposite 10 to show the pupil's success on that day. Then a line is drawn between these two circles, thus making the first section of the curve. The rest of it is built up in similar manner. Notice that the curve rises with gains and falls with losses.

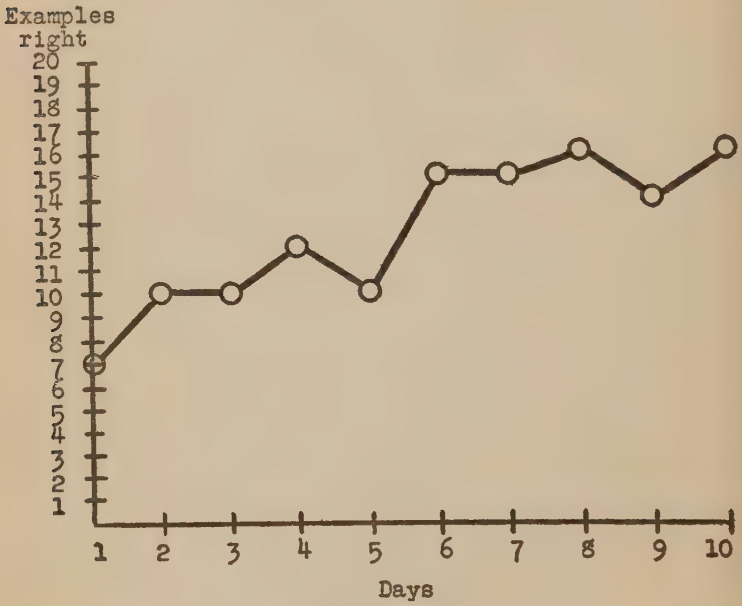

FIGURE 2.

This sort of curve is called a learning curve, because its form shows how fast the person in question learned during his successive periods of practice or drill.

Learning of most kinds takes place rapidly at first, then more slowly until practically no progress is made unless new incentives or better methods are employed. During the periods of no progress the learning curve tends to be flat or even to drop a little. In Figure 2, for example, progress is good up to the sixth day, and after that it almost ceases—day-to-day fluctuations being disregarded. If after the first flattening of the curve has taken place, the learner is induced to put forth greater effort or use a better learning technique, the curve again rises for a time and then flattens out again. Still further incentives, interests, or efforts, and still more effective methods of learning, may cause the curve to rise successively to new and still higher levels.

These flat or level sections of the learning curve—corresponding to periods of no measurable improvement—are called *plateaus;* and they are characteristic of all learning that is carried out over a relatively long period of time. If we learn typewriting, or billiards, or French, we usually make rapid and encouraging progress at first. Soon, however, follow the drab days when we gain nothing. Interest flags and the temptation to quit is ever present.

Here the teacher's function is important—far more important in helping forward the learning process than is usually supposed. In the first place, the teacher must be keen to detect the onset of these periods of arrest. The pupils should then be shown that such periods are practically inevitable, that everybody has them, that they (the pupils) are really making progress—as indeed they are—only it cannot be seen, that they are organizing their experience, establishing their former gains on a more secure basis, and making ready for a new advance.

Experimental data sustain this view. In typewriting, for example, it was found by Book that improvement is first made on the basis of pressing each key as a separate act. Soon the learner has made all the progress he can make on that basis. He reaches a plateau, a period of no improvement. But if he persists, he will gradually work out a new procedure. The letters that he formerly typed separately will begin to flow together, first the more usual combinations such as t-h-e and i-n-g, later the combinations that occur less frequently. These will begin to run off together and to be written in each case with one nervous impulse. Then the learner will leave his plateau, and his learning curve will begin to rise again.

Presently a second plateau will be encountered and again, if further progress is to take place, reorganization into more efficient habits must occur—habits in virtue of which longer words and whole phrases will be written together as one act. If this reorganization is accomplished, the learner will enter upon another period of improvement; and each time he persists through the discouragement of a plateau, he will rise to new heights of proficiency and expertness.

This is verily the price of success in any line of endeavor—persistence, dogged tenacity in the face of discouragement. If a person—you or I or our pupils—can learn this and put it into practice, he need have no concern for the future. As James says in his ever famous chapter on habit, "He can with perfect certainty count on waking up, some fine morning, to find himself one of the competent ones of his generation, in whatever pursuit he may have singled out."

And so I say that the wise teacher will, in the first place, detect the onset of these ominous plateaus and will in an elementary way acquaint the children with their meaning. In the second place, the teacher who knows these important facts about learning will apply extra incentives, appeal to new interests, and suggest more potent learning methods in order to carry the pupils over these uninspiring times of apparent arrest.

THE LAW OF EFFECT

In the beginning of this chapter I asserted that the law of effect—to use Thorndike's terminology and analysis—is fundamental in learning. In accordance with this law, if something we do proves to be satisfying we tend to repeat it or to continue it as long as it continues to satisfy us; but if what we do proves to be annoying we tend not to repeat the activity or to discontinue it. This is the emotional side of learning. Here belong interest on the part of the pupil and motivation as a duty of the teacher. Games, applications of knowledge or skill, projects, the sense of achievement—these are some of the positive aspects of the application of the law of effect. If the rôle of the teacher includes, as I have pointed out, the preparation of the pupil for learning and the provision of right conditions for the learning process, thus serving the law of readiness, it is no less the teacher's task to keep interest alive and to sustain the pupil throughout his learning upon the level of his best effort. But particularly at the beginning of a task and again at each of those critical stages where interest tends to flag and the weak and misguided lose heart—at these times the kindly offices of the teacher are most needed.

One type of application of the law of effect may be cited as typical of the favorable influence which it may be expected to have when properly applied. One of the most potent incentives to effort has been found to be a clear understanding and appreciation on the part of the pupil of the objective he is driving at and a knowledge of his success in approaching it. Certain investigators—notably Judd and Arps—have driven home the idea that practice with knowledge on the part of the learner of the results of his practice is far more productive than the same activity without knowledge of results. Practice without such knowledge is blind. The pupil cannot organize his efforts for new conquests, because he does not know whether the efforts he has already made have been successful or unsuccessful. Consequently the devices and methods he adopts today he does not know whether to abandon or develop tomorrow.

One means of conveying to pupils a knowledge of their success is to have them *plot their own curves of learning*. This has been found to be advantageous with children in grades as low as the third. For example, in O'Brien's experiment in improving the rate of reading, some sixteen hundred children in grades from the third to the eighth kept their progress records in chart form for a period of thirty-eight days. It was the unanimous testimony of the teachers of these children that this was a powerful device in motivating their work and thus in securing sustained effort. No child failed to grasp the idea that when his curve rose he was succeeding and that when it fell he was not.

The statements I am making will, I judge, commend themselves favorably to your belief. The literature of psychology and of education is replete with data on the favorable and unfavorable influences of the emotional accompaniment of learning. For example, Gilchrist reporting on the "extent to which praise and reproof affect a pupil's work" shows that when the persons who took his first test were told they had done well, they improved 79 per cent on the next test.[22] On the other hand, he shows that those who were told they had done poorly did not do so well on the second test as they did on the first.

The task for the teacher is to banish the unpleasant and therefore unfavorable feelings—the "annoyers," as Thorndike calls them—and to bring into play the pleasant and therefore favorable feelings—the "satisfiers." I cannot tell you in quantitative terms how many annoyers and satisfiers it takes to produce a given decrease or increase in the results of learning; and no unit amount of satisfaction or of interest has ever been devised. The experimental work, therefore, which can be brought under the topic the "law of effect" is restricted. It would, however, be the grossest error to suppose that because its treatment in this connection is short, the law of effect is of secondary importance. All the other laws and principles which have been described in the foregoing account are inoperative without it. Thorndike has said somewhere that it is not a question whether we shall teach with or without interest; we cannot, in fact, teach anybody anything unless he is interested.

---

[22] *School and Society*, 4:872-74 (December 2, 1916).

Although, as I have said, the amount of experimental work under the heading of the "law of effect" is restricted because we cannot measure satisfiers and annoyers, it would be a mistake to suppose that experimental work cannot be undertaken. You ought to try out the value of the learning curve as an incentive, especially when each pupil makes his own. You can find out the extent to which, in your own class, work without knowledge of progress is blind. You can utilize different incentives when interest flags and plateaus are reached, recording at least your impressions of their value. You can even measure the value of any given device or procedure by a "before-and-after" testing. You can verify Gilchrist's experiment; and if you do you will obtain a more vivid realization than ever before of the bad effects of a cross, nagging, scolding, or otherwise disagreeable manner.

SUMMARY

In the foregoing treatment of learning I have regarded education as adjustment, the bringing about of desirable changes in the student. As such it is a complex process requiring the services of a professional class.

I have asserted that teaching should, and indeed must if it is effective, parallel learning. As Trabue says, it is as idle to talk about teaching when no learning is taking place as it is to talk about selling corn when no one is buying corn. "In the future," as Rusk says, "we shall . . . have to talk less of the teaching process and more of the learning process, and for guidance in method we shall have to depend upon the psychology of learning instead of on 'formal types' and the logical analyses of knowledge." Accordingly I have maintained that a knowledge of *how learning takes place* is indispensable for teachers.

The distinction between passive and active learning is made and, with reference to the former, the rather neglected potency of right environmental conditions is pointed out.

The remainder of the foregoing treatment may be broadly subsumed under three headings—to which, however, I have given treatment of exceedingly unequal length—namely, the laws of readiness, of exercise, and of effect. These laws, however, before they can be applied need to be analyzed into subordinate laws or principles and these in turn into specific activities.

The law of readiness asserts the fact that before learning can take place the organism must be ready for it. The teacher's rôle is to know each child's abilities, to make sure that he is ready for the learning he is required to do, and to provide favorable conditions and attitudes.

The law of effect is concerned with the emotional accompaniment of learning. It is closely connected in the positive sense with interest and negatively with feelings of distaste or annoyance. The rôle of the teacher here is to maintain the morale of the classroom and to sustain pupils on the level of their best efforts.

The law of exercise has to do with the repetition which is so generally necessary in learning. Since the experimental work connected with learning falls naturally under the law of exercise, I have given a rather extensive discussion of certain groups of experimental data.

Periods of practice should be short and frequent unless other conditions than economy of learning are to be considered. This fact is well established; but it is also asserted that the periods can be made too short. The teacher has an excellent opportunity for research work to determine for his class and for individual pupils what the "best" length of period is.

Some learning, however, involves little or no repetition. This is especially true of observational learning and of learning in which a high degree of association is present.

In the types of learning which require repetition—and most learning does require some repetition—the question of how the repetition shall be made is important. This gives rise to a discussion of the whole and part methods of learning. It appears that the natural method is not the economical one. The evidence is abundant that untrained persons, both adults and children, when asked to memorize a given piece of material, break it up into small parts, memorize these parts, and then put them together. The evidence also is to the effect that this part method is not in general as economical as the whole method. This is true for immediate memory and is even more strikingly the case for permanent retention. It appears also that the conditions which favor pronounced superiority for the whole method are: good ability on the part of the learner, uniform difficulty of material, and considerable length of material.

While the advantages lie with the whole method as opposed to the part method, a mixed method is better than either. According to such a method the learner should go over the material in the beginning, select especially difficult or important parts for immediate learning, then divide the entire material into sections, surveying the whole from beginning to end and pausing at the end of each section. The movement, however, should be progressive from the beginning to the end of the material. The pauses are designed to permit a freshening of attention.

Under the heading of "forgetting" I have shown that rapid learners retain best; that people differ less as to their forgetting than as to their learning; that each person has, at least for definite materials, a learning index; that memory may be improved in the field in which it is exercised; and that relearning sufficient to offset forgetting may be provided at intervals which apparently obey a law of progression.

Many investigators have observed the opposite of forgetting, namely, reminiscence—the improvement of a function between periods of practice. More information about this phenomenon is needed in order that advantage may be taken of it. In securing this information the classroom teacher can be of service.

I have taken the position that transfer of training is real and that more attention should be given to its practical manifestations and less to controversies about it. In practical work the teacher will find that two factors operate. One is the transfer from old knowledge, habits, skills, ideals, and attitudes to the new situation. The other is the new element which must be learned. The extent to which transfer may be relied upon and the consequent extent to which new learning must take place are matters for the exercise of fine judgment on the part of the teacher. On the one hand, he must not teach everything as if it were new; and on the other hand, he must not trust to a total transfer from one situation to another.

It is my belief that the doctrine of types of learning is useful in school work. I have myself seen, and others have reported, misunderstood children because their type of learning was not the type required by the school. The report by Fernald and Keller is cited in this connection. Rough but reasonably adequate means are available to the teacher for determining, especially as to the few problem cases in his class, whether or not the children are being forced to employ modes of learning to which they are not naturally adapted.

It does not appear that mental fatigue as distinguished from mere feelings of fatigue need give the teacher much concern. The evidence is that within the ordinary school day the pupil who is in normal condition may work about as well at one time as at another. The mental hygiene of the schoolroom should concern itself with more and better incentives and a pleasurable emotional tone.

I have briefly discussed the nature of the learning curve. I have pointed out the prevalence and meaning of plateaus and the rôle of the teacher in connection with them.

The law of effect, in virtue of which satisfying activities are prolonged or repeated and annoying ones are not repeated or are discontinued operates throughout all learning. It is closely related to interest, and no technique either of teaching or learning having to do with the distribution, the number, or character of repetitions will much avail unless the work is interesting.

Throughout the foregoing chapter I have asserted the privilege and the opportunity which the teacher has of engaging in research work even in matters such as problems in learning which have been hitherto exclusively the business of the psychologist. In the first place, many of the more important conclusions of the psychologist—very often determined from the study of adults—need confirmation with children in the classroom. In the second place, there are important problems which the teacher is in a favorable position to undertake from the beginning. In the third place, even among the truths which have been undeniably established, the repetition of the experimental procedure by the teacher will make the facts more vivid, convincing, and meaningful. Finally, the general solutions of the psychologist need special application to particular pupils; for to most questions the true answer is not *one* but *many*. There may be one answer applicable to average conditions and to pupils in general, but teachers deal with particular conditions and with individuals. The general answer, which psychology often attempts to supply, has in some instances little bearing. It requires modification this way or that according to the conditions; and these conditions are much more numerous and divergent than any psychological laboratory can set up. Their effects can only be determined when a great many learners are available for observation—learners using various types of procedure, and exhibiting different stages of maturity, levels of intelligence, and degrees of effort—learners working on various kinds of tasks under various incentives and with various purposes in view.

The psychological laboratory has no such resources at its command. It commonly works with but a few subjects. Ebbinghaus, the German psychologist, wrote a great book on memory, but he based it on results secured from but one person. Binet, the well-known French psychologist, wrote an important book on intelligence, but he secured the material for it from but two persons. It naturally follows from the fact that the psychologist has but a few learners to observe that he can pursue his investigations only in narrowly prescribed fields. The massing of homogeneous data, the comparison of types with types, the checking up of hypotheses—these must be reserved for a larger theater, namely, *the school.* The psychologist will himself enter the school, and not the least of his service there will be the training and the inspiring of teachers as research workers.

CHAPTER II

WHAT STATISTICS DOES THE MODERN TEACHER NEED?

ALL of us who are engaged in education hold certain opinions. We make assumptions as to the worth of certain items of the curriculum and certain ways of doing things. We have certain convictions about education itself. For example, one of my beliefs is that the schools of today are better than they ever were; but there are some who do not agree with this idea. If I am to make my case, I must find some means of measuring the work not only of the teachers of the present but also of those of the past. I must also have recourse to some method of handling the measurements after securing them, so as to bring out the comparison to which attention must be drawn. In other words, I shall need to invoke the aid of statistics.

OLD SCHOOLS AND NEW

Caldwell and Courtis have done this in their book *Then and Now in Education.*[1] This book takes us back to the days of 1845 in the city of Boston. Horace Mann was then pointing out the shortcomings of the schools of that city—shortcomings in comparison with his ideals rather than in comparison with the schools of other lands or of earlier times. The Boston schoolmasters retorted vigorously. At length the committee, whose duty it was to visit and report upon the schools annually, conceived the idea of accomplishing their purpose by means of a carefully constructed and uniformly applied battery of written examinations. They gave these examinations to pupils who were approximately fourteen years old and who were selected by the masters to take the examinations. Fortunately the examining committee reported not only the questions asked, but also the number and per cent of correct answers to each question and even the kind of answers which were regarded as correct or incorrect. Luckily, too, a copy of this report has been preserved to the present time; and not long ago it fell to the good fortune of Caldwell and Courtis, the authors of this book, to give the same tests to children of the same age in the schools of today.

---

[1] Caldwell, Otis W., and Courtis, Stuart A. *Then and Now in Education.* Yonkers, New York: World Book Company, 1924. 400 pp.

The obviously interesting problem is the comparison of the modern school with that of 1845. Although the authors tell us at the outset that the Boston schools were better than the schools in most parts of the country, although they quote Horace Mann to the effect that Boston was then spending more money on its public schools than the Parliament of Great Britain was spending on public education in all England, although—on the other side of the picture—they tell us of the lack of organization, of the mean buildings, and of the perpetual floggings by which the schools were then characterized—although the authors do these and similar things, the nub of the problem still remains untouched.

Despite their shortcomings the "good old days" still hold their glamor. Does not the business man tell us that children cannot add and spell as well as they used to? Is not the halo of the little red schoolhouse still visible? And were not the Boston schools the best in the land in those days? It is difficult to believe that these children of 1845, representing as they did a selected group from homes where education was appreciated, would prove to be inferior in point of scholarship to the unselected masses of the pupils of today. Moreover, we may have in mind the fact that the examination questions propounded seventy-five years ago would in many instances be entirely inappropriate today. In arithmetic, these were the days of duo-decimals, of allegation medial, of compound proportion, and of complex fractions. In spelling, a sample page from a textbook used in Boston in 1845 contains such words as *circumambient*, *risibility*, *Bacchanalian*, *pericranium*, and *contiguity*. In grammar, it was the day of parsing, and in geography, of memorizing capes, peninsulas, and bays. Yet Caldwell and Courtis by their method of handling the data are able to show that on a fair and comparable basis the children of 1919, even when unselected, made a better showing than did the selected children of 1845 on test questions designed expressly for the children of the earlier period.[2]

But it would have been quite impossible for Caldwell and Courtis to show this if they had not had at their disposal a certain statistical technique. Without a means of measurement alike applicable in 1845 and in 1919 and without a method of combining and comparing in numerical terms, they could not have made their case in the mind of anyone who was disposed to take the opposite view.

---

[2] Of course the questions on subject-matter entirely outside of the course today—for example on allegation—were omitted. The questions, however, that were used were the same as were used in 1845.

PROGRESS AND THE NEED OF MORE EXACT KNOWLEDGE

Not long ago the question of whether one method of teaching or one type of school organization was better than another would have been decided, just as the Boston schoolmasters wanted to decide the issue with Horace Mann, by vociferation and abuse. As long as we have no units upon which we can rely and no means of handling them which commands confidence, we continue with poetic vagueness to describe a large number as being like the sands of the sea or the stars of the heavens.

As men have come to think more accurately and as the demands of modern society have become more exacting, we have been forced to adopt more satisfactory methods of expressing our thoughts, of describing phenomena, and of forecasting probabilities. With measurement and the statistical method, science becomes possible. Alchemy becomes chemistry, and astrology becomes astronomy. The compass, an instrument for measuring direction, and the thermometer, an instrument for measuring heat, satisfy the requirements of scientific accuracy.

It is not remarkable, therefore, that during the past fifty years—years in which the American people have committed themselves more and more to the policy of education at public expense—the same development has taken place in school work. It has been increasingly required of teachers that they render an account of themselves. With the advent of the principal as a supervisory official and of the superintendent as both supervisor and administrator, the requirements of the public for information and the technique by which this information can be supplied have developed together.

To such an extent has this been true that there is coming into being in the larger and more progressive city systems a new type of expert service. This is performed by one who is variously called the efficiency man or the director of research or the statistician. But despite this specialization there yet remains for the teacher a far larger amount of statistical work than has ever been the case in the past. There are several reasons why this is true.

STATISTICS AND RECORD FORMS

In the first place, if the superintendent is to make a report he must obtain data for it, and these data come in part from the teachers. Unless the teachers understand the statistical meaning of the items in their register and other record forms, they are likely to record fundamental information inadequately. Having little idea of the intrinsic value of these data and still less of their importance in the more general statements and reports relative to the work of the schools, they are inclined to belittle the importance of this part of their duties as teachers. I have always thought that the method which obtained some time ago—and perhaps is still in effect—in London, whereby teachers-in-training were required to keep, under close supervision, all the forms used in the public schools, was a particularly sensible arrangement. Teachers are not born with a sense of statistical values, and many of them keep even the simplest records and make out the most obvious reports with astonishing inaccuracy.

Just now I am going through some data which I have received directly from teachers. One item reads: "No. of years of teaching experience in Ohio: . . . ." A certain per cent of the teachers responding to this item write "Yes" or "No." Again, in calling for a description of the position held by the teacher, the first item is: "Name of city (or district and county)." Some, instead of giving the name of the city or of the district and county, write "City" if they are teaching in a city, or "District and county" if they are teaching in a rural district. Perhaps you will say that these headings are not altogether clear. But I think that if the statistical values of the replies were remotely apprehended, one would not need a heading which forestalled every possible understanding except the one intended. Moreover, if such headings are used—I mean headings which preclude misunderstanding even by a very ill-informed or careless person—they become involved and clumsy. They take on many qualifying expressions, each designed to head off a particular kind of error. The result is that the heading when thus made "fool-proof" becomes hard to understand for another reason, namely, because of its complexity. Something should be expected from the teacher, *some general notions regarding statistics and the record-making which lies at the basis of statistics.*

The following story will still more accurately illustrate my meaning. On a certain occasion I obtained from about seventeen thousand teachers a few simple facts which they recorded for me on 5″ x 3″ cards. The first item on each card was "Name ...." On the back of the card directions were given for each of the four or five items on the front. The direction pertaining to the first item was: "Write your last name first." My recollection is that about three hundred cards came in with the given name written first. At any rate about one in fifty either misunderstood or neglected the direction. Since I desired to file the cards in an alphabetical arrangement according to the surnames, and since I wished to have the entries in the teachers' handwriting, I returned the cards to them, calling attention to the direction which they had failed to follow. All the cards came back made out as I had intended except one. Sarah Jones had written her name just as she was accustomed to sign it. Being somewhat interested in Sarah, I sent for her. Upon her arrival I placed one of the cards before her and asked her to read the direction for the first item.

She read it: "Write your last name first."

"Now do it," said I.

Taking one of the cards she wrote *Jones* toward the right of the space provided for the name and then wrote *Sarah* in front of *Jones*. *She had written her last name first;* and in the discussion which followed she had entirely the best of the argument. I said that what I wanted was the last name in the first position, in order that I might file the cards alphabetically according to the last name. The substance of her reply was the gentle query: "Why didn't you say so?"

Yet the mentality of an adult who has arrived at the necessary training and culture to teach children might be expected to reject an idea so grotesque as that anyone would care whether or not she wrote her last name first in point of time. If she had possessed any notion of the way statistical data are handled, such a notion would have saved her from so naïve a blunder. The modern teacher needs to understand the statistical meaning, importance, and use of the items which usually appear on educational record forms.

In the second place, so rapidly have means of measuring the products of instruction developed and so widely have they been applied, reported, and discussed that they have carried with them far more frequent and exacting numerical treatments than was the case before these tests and scales were devised. The classroom teacher, especially one in a progressive environment, can scarcely escape the requirement that he give standardized tests and render reports concerning them. These tests are usually accompanied by score-sheets, yet many teachers are unable to understand them. Unless a principal or superintendent can obtain these score-sheets correctly made out, he cannot accurately report the achievement of the pupils in the various rooms and grades.

Likewise, as soon as two tests—say, an arithmetic test and a reading test—are given, the natural question arises, "Did the children do better in arithmetic than they did in reading or vice versa?" Moreover, a measure of the difference, if any, is desired. Suppose a child scores 7.9 in a certain problem test in arithmetic and 21 in comprehension of reading according to a certain reading test. Which score is the better? Twenty-one *looks* better than 7.9, but if 21 turns out to be the reading standard for the fifth grade and 7.9 turns out to be the arithmetic standard for the seventh grade, the meaning of the figures is decidedly not the apparent meaning. In order to interpret these measures, one must reduce them to a common unit, and the ways in which this may be done involve a statistical treatment which most teachers do not possess.

In the third place, the literature and the platform discussion of educational questions during the past ten or twelve years have been completely transformed. They are shot through and through with statements in the form of numbers. Although general ideas and rhetoric met all the requirements of an earlier day, the present period is more exacting. This period requires an author or a speaker to tell how many and how much, to compare and relate, to give expressions of typical or average conditions, to show the extent to which the phenomena vary about this typical or average condition. Accordingly, in order to understand such writing and speaking—in order to be abreast of the times—the modern teacher must know something about *statistics and statistical method.*

Thus on several accounts, for the teacher scarcely less than for the administrator, statistics has come upon the scene with its new and disquieting jargon and its strange tabular and graphical displays. How much of this does the modern teacher need? Much obviously depends upon what the modern teacher is trying to do. Much, too, depends upon what we understand statistics to include.

I distinguish the modern teacher from the teacher of twenty-five or thirty years ago as a person who desires to have at his command the results of that surprising period of research which entered upon its present stage about the year 1910. It is true that many of these results are not easily accessible, but part of them have been worked over into textbooks and curriculums and a still greater part into books on method. The modern teacher utilizes as much of the product of research as he can, but the amount he uses is perhaps less important than the spirit he displays. The modern teacher is, above all things, a student of his calling, and the aim of his studentship is expertness in his calling. For the present, therefore, we may consider that teacher as modern who is reasonably in the possession of the principles and technique of teaching and who possesses what I have called studentship in reference to his work.

For the purpose in hand I take it that statistics may be made to include all the numerical treatment of educational facts for which this modern teacher has, or ought to have, a direct use. This will include the record-making which lies at the basis of statistical technique. It will involve the ability to understand the terminology of educational reporting, the ability to record the facts of his experience according to this terminology, and the ability to work up the data into usable form.

The ability to understand the terminology of reporting will involve an appreciation of that peculiar form of expression known as *precis* writing—the concise system of headings and subheadings which characterizes report forms. This ability is by no means to be assumed, as an examination of the register and other blanks kept by teachers in almost any school will quickly reveal.

Consider, for example, the following very simple form.[3]

---

[3] Strayer, G. D., and Engelhardt, N. L. *The Classroom Teacher.* New York: American Book Company, 1920. p. 306.

HORACE MANN ELEMENTARY SCHOOL

Parents' Report

1918-1919 ....*Quarter*

Name........................ *Grade*........ *Room*........

I. Studies
  1. Above average....
  2. Below average....
  3. That show particular improvement....
  4. In which child apparently is not working to the best of his ability....

II. Habits and Attitudes Desirable for Good Citizenship
  1. Which are unusually well developed....
  2. Which need careful training....
  3. In which improvement has been marked....

Number of days in quarter...Days absent...Times tardy...

The correct understanding of this form involves an appreciation of the extent to which each heading is effective. In the first place, the heading "Horace Mann Elementary School" should indicate that nothing about studies, citizenship, absence, or attitudes has reference to these as exhibited in any other school except this one whose name is at the top of the form. Furthermore, the figures "1918-1919" have a similar limiting influence. In other words, their position tells us that they carry their meaning throughout the entire form. They also qualify or limit the meaning of the name of the school. They tell us, in other words, that data regarding studies, habits of citizenship, etc., are not wanted for the pupil in question throughout all his course at the school but only for the year 1918-1919. It would therefore be a breach of good reporting for a teacher, even though he might know something regarding these matters as they had revealed themselves the previous year, to include any matter pertaining to the earlier year.

There is a further time limitation, namely to a particular *quarter* of the year 1918-1919. The teacher, however, is expected to indicate the period to which the report refers by writing a word or number (the directions should tell which) in front of the word "quarter." When this has been done, nothing that follows has reference to any other quarter.

The next line of the report [4] introduces a very sharp limitation. It is apparent that the report is not intended to show facts regarding a class or any other group of pupils, but that whatever facts it contains have reference to a particular child. Moreover, from the form of the headings on this line the rules of reporting will tell the teacher how to proceed. The word "name" is a caption which, when taken in conjunction with the space following it, directs the teacher to write the pupil's name immediately after this caption. Similarly the captions "grade" and "room" with the spaces which follow them direct the teacher to make appropriate entries. These entries, it must however be understood, are governed by the pupil's name which has been written immediately before them. In other words, the grade must be the grade of this pupil as it was in the quarter which has been designated and for the year 1918-1919 in the Horace Mann Elementary School. The teacher must not enter the grade of some other pupil or the grade of this pupil in a different year or quarter or in a different school. The same

[4] I leave out the words "parents' report." This is merely a title which serves to distinguish this form from some other form; though it is true that even this term has a limiting meaning.

may be said as to the entry after the caption "room." The influence of previous headings must be kept in mind.

In order to continue with the proper making of this report the teacher must next understand the meaning of the "hanging" and indentation by means of which general and subordinate headings are conventionally represented. He must know that the four indented headings refer to "studies" and to nothing else. Thus, after the words "above average" are to be written the *studies* in which this pupil—while in the room and grade indicated, during the quarter and year indicated, and at the Horace Mann Elementary School—was above average. Similar statements may be made with respect to the other three indented items.

In the next place it is to be observed that the heading "Habits and Attitudes Desirable for Good Citizenship" stands in the same order of importance as the heading "studies." In other words, it is not in any sense subsumed under "studies" nor in any way limited by it. It goes back for its limitation to the same set of headings as does "studies" itself. On the other hand, the three indented items which follow refer only to *habits and attitudes desirable for good citizenship*. The teacher is called upon to name first, those "which are unusually well developed"; second, those "which need careful training"; and third, those "in which improvement has been marked."

Finally, a series of three items, each involving a single numerical entry, show by their style of type that they are coördinate with "studies" and "habits and attitudes desirable for good citizenship." The "number of days in quarter" must refer to the same quarter as is indicated above and, if this number can vary in different grades and rooms, the entry must be for the grade and room already set down.[5] Similar limitations exist in the case of "days absent" and "times tardy."

All this may seem very obvious, but long experience with the vagaries shown by people in making out report forms assures me that it is not. A neglect of some of these apparently obvious principles of subsumption and limitation may be exhibited in the reporting even of those who are supposed to be specially trained for it. A statistical clerk in one of the state departments of education lately made out a blank calling for certain financial data. A general heading read as follows: "Expenditure of the department for itself." Under this and indented, as the subitems are in the above form for the Horace Mann Elementary School, were three items and a fourth item called "total." This total meant—and this would be obvious to anyone who knew the language of reporting forms—the total "expenditure of the department for itself." In other words, it was governed by the caption under which it was subsumed. Nevertheless the statistical clerk to whom I refer entered opposite this total the entire expenditure of the department including state aid to weak districts and other subventions which were not in any way for the work of the department itself.

In a form which I have lately been using and to which I have already referred, the following set-up occurs. It relates to the position which the teacher in question holds.

---

[5] Technically, this item should have been placed *above* the line on which the name is entered. Can you tell why?

Location	
Name of City (or District and County)	*State*

Now, anyone who knows the meaning of the placing of items and subitems does not need to be told that what is wanted is the location of the position to the extent of naming the city and the state or naming the district and county together with the state. I have already mentioned the form of error which occurred scores of times. Under the caption *Name of city (or district and county)* was frequently written the word "city" or the words "district and county."

I shall not dwell longer on this matter of the ability to understand record and report forms. I grant readily that many of these forms are faultily constructed; that they do not convey the impression which their authors intended they should. But on the other hand, a great many of these forms are properly constructed, and the number of such is increasing every year. We now have whole systems of report and record forms which have been carefully devised by those who know how to do such work. It seems to me that teachers should be taught while they are in training to make out such forms. I know from experience that ability in it may not be assumed. If it is really—as I think I have shown—part of the expected duty of teachers, then it must be provided for. It is my advice to those who are now in service and who did not have instruction of this sort, that they obtain samples of forms such as the Strayer-Engelhardt cards and study them.

FILING AND INDEXING

The statistics which the modern teacher needs involve the classification of information. One method by which this classification may be achieved is by the actual arrangement of the material itself. This leads me to a consideration of filing and indexing as it pertains to the work of the teacher.

You will find many books on filing and indexing, and briefer statements of the subject in books devoted to office practice and business procedure. The material with which these books deal relates exclusively to commercial and industrial matters. The paper to be filed consists of letters, orders, vouchers, credit data, and all the various materials which are found necessary in conducting an established business. The time has passed when anyone tries to remember the location of a letter or the details of a transaction. Business methods have passed rapidly from the days when the wire spindle was thought to be a sufficient filing system to the days of vertical files and card indexes.

Teaching is by no means a commercial enterprise, and yet the body of material connected with it is far too great to be retained in memory and much too valuable to be thrown away. If an act on the part of the teacher when once performed was no longer of any use, one would be justified in forgetting it and in failing to make a record of it. But such is frequently not the case. Each act of teaching or of classroom management, each item of information about the children of a given class, every plan made, and every item of teaching material collected, is as likely to have bearing upon future action as the correspondence which a business firm files so carefully.

One of the great defects in teaching is the failure of teachers to profit by their own experience. The professional training which a teacher obtains in an institution prior to securing a contract to teach is less important in determining the expertness of the teacher than the way he uses his time on the job. Of course we have here the operation of that enormous waste due to the short tenure of teachers. But what I have in mind is the teacher who has been in service, say, five or more years. Such a teacher should have acquired personal characteristics and modes of procedure far more effective than those which he had during the first year of service. He has the capacity for growth from the beginning of his service and that capacity should be brought into vigorous play and made to serve throughout his entire career. It is believed by some who have investigated the question that during the first five or at most ten years of service a teacher acquires all the competence which he is going to acquire as the result of experience. There is no reason why this should be true. There is no more reason why a teacher at the age of twenty-five or thirty should stop growing professionally than there is that he should cease to develop in personality. At that age the teacher has not much more than begun to reach maturity.

The outward and visible accompaniment of continued growth in expertness may reasonably be expected to be some material possession which distinguishes the experienced from the inexperienced teacher. The experienced teacher may be expected to have more and better tools at hand. Even during the first year of service the teacher should begin with a small filing system. An ordinary box file which can be purchased for fifty cents may serve at the beginning. There is no reason, for example, why the test you gave last week in fourth-grade arithmetic should be thrown away after it has served its immediate purpose. You ought to keep it and indicate on it at least the grade to which it refers—do not try to remember even this simple fact—and the date at which it was used.

Teachers are continually devising tests on the spur of the moment. Sometimes they make them up as they go along; and they are often quite unaware of the inadequacy of tests made in this manner. In arithmetic, for example, one of the clearest results of modern analysis is the fact that textbooks and examinations do not distribute the requirements for practice or testing over anything like the entire field with which they are concerned. They run in grooves; and the same number facts and the same situations occur again and again, while other equally important facts and situations do not occur at all. When I construct a test in column addition, I have a special fondness for the number 217. There is no reason for it that I can discover, but it is a fact. It is certain, too, that I have other predispositions of which I am unaware; and unless I adopt some plan for distributing these number facts, it is inevitable that I shall omit a great many of them.

Keep your test papers, then, and label them so that their purpose is permanently recorded. It will also be well to file with them at least a few of the more interesting answer papers submitted by the pupils, for example, a few of the best, average, and poorest papers, and the papers showing important errors. If your files thus permit you to analyze your test papers and the pupils' answers in a given subject to see, on the one hand, what you really did require and on the other hand what the typical errors of the pupils were, you will be adding greatly to your expertness in giving tests, and better still, in teaching your pupils so as to meet their needs.

Of course I am using the idea of filing test papers merely as an illustration. Every teacher worthy of the name should plan his work daily. Only geniuses and fools work without a plan. These plans should be filed; and at the time of filing, a notation should be made indicating the extent to which they were carried out. Material of this sort will be of great value if properly used in subsequent years. It will enrich your experience. It will give you two years of growth-on-the-job in one year of time. It will prolong your years of growth.

I know that in school systems where plan books and progress books are required by the administrative authorities, teachers quite generally look upon them as something done merely to satisfy a demand from above. In that frame of mind they seldom make actual use of the plans for the benefit of their own teaching. The plans and the progress records are things to be made rather than things to be used. They are to satisfy someone else rather than to satisfy oneself. Done in this way they have little if any value. Yet it is certain that if these plans and these records of progress become *your* plans and records made under the belief that your growth as a teacher will be served by them, then they become real instruments in your development.

Miss Emily H. Welch in an article in the *Elementary School Journal* for February, 1917, suggests the use of progress books as an important aid to the teacher. "In this day when accurate, complete, and permanent records of financial expenditure, of equipment and supplies, of attendance, and of achievement are considered to be the *sine qua non* of any well-organized school office, a discussion of other phases of this administrative problem should be of permanent interest." The progress book is suggested as a record of the work actually done from day to day and from week to week in each class.

Every Friday the teacher writes up the work of the week. For each subject he uses a separate sheet, perforated for loose-leaf binding. His statement includes the general topics that have been discussed, the number of pages of the textbook covered, any experimental work done in a scientific subject, the list of problems in mathematics, and any other data that need to be recorded. Brevity and conciseness are important considerations.

At the end of the year these sheets are bound together and put in a stiff cover along with the following additional information: (1) a list of the members of the class with a note of any irregular entrance or withdrawal, (2) a list of the textbooks used and a note of portions omitted, (3) a condensed account of the ground covered by the course, and (4) the name of the teacher.

Thus a teacher may check up his own work by comparing his weekly progress in successive years. He may also acquire a better perspective of the year's work as a whole. Two teachers teaching the same subject may use the progress books as a basis for comparison and discussion. A new teacher can become thoroughly familiar with the material used and the methods employed by his predecessor.

These books are perhaps of the greatest value to the principal, for in them he has in a compact, concise form a complete story of the academic work for the entire year, and during the year he can watch the progress of the work in each class and can know the point reached at any moment.

Again, as the teacher acquires experience he should gather materials for use in teaching. One of the things which should distinguish a teacher who has taught five years from one who has taught but one year is the fact that the former owns more and better tools for his work. Suppose, for example, that you are teaching fifth-grade geography for the first time; and that this means the teaching of the geography of the United States. If you pass through that period of experience without acquiring, not only for your present class but for future classes, an equipment of stories, articles, pictures, charts, and specimens pertaining to at least some features of the geography of the United States, then you have failed to render to yourself a good accounting of the time you have spent.

As you pass through the various experiences which become your privilege as a teacher, you should be alert to collect and to preserve all sorts of interesting and effective teaching materials. This applies to every subject and to every level of teaching—to history, mathematics, and science in the high school and to spelling, silent reading, and story telling in the primary grades. Teachers of arithmetic commonly complain that the textbooks do not contain enough problems. They do not, and they probably never will. The remedy, however, is simple. In addition to the textbooks' broad guess at the needs of each grade, the teacher should make up with care and secure from the pupils and the community problems and examples which make a more accurate adjustment to immediate local needs. And this material should be preserved. It should be carefully made up in the beginning and it should, therefore, be worth preserving. Moreover, it should be preserved in such a way as to be readily consulted and applied to future conditions and needs.

This means filing; and there is as legitimate a demand for it in teaching as in business. The teacher who starts with the determination of profiting by his experience will soon pass from the period of the box file to that of the vertical file with its accompanying card index. I was about to say that every schoolroom should have a vertical filing cabinet and a small cabinet for 5″ x 3″ cards. I will not say it, not because it is not true but because the teachers in thousands of rooms would make so little use of this equipment that the expense of its installation would not be justified. Perhaps the better way to put it would be to say that every room should have a teacher who was able and willing to use such equipment.

It is not my intention to enter in any extended way into the different types of filing which would be proper in a classroom. Perhaps a chapter or even a book should be devoted to this question, but it does not fit into my present purpose. I shall content myself with pointing out that material such as clippings, pictures, test papers, etc., should be filed topically and in alphabetical order; and that this material should be indexed on cards according to the plan which will permit the most effective use of it. Thus a card is made for each picture, article, clipping, or test paper.

The idea is this: The material itself can only be filed in one way—unless you have duplicates of it, which is not usually the case. But one frequently wishes to refer to it from a point of view different from that used when filing it. Accordingly, each piece of material has one or more index cards as its substitutes or representatives. An index card may be put in a different arrangement from that in which the piece it represents is filed. This will often permit one to find a given piece, when to do so without the aid of an index card would be all but impossible.

For example, one might file test papers and pupils' answer papers under a subject heading—the Revolutionary War, or Long Division, or Great Britain. But there may also be need for locating a particular pupil's paper or a paper exhibiting certain types of error. In this case, if index cards are made and arranged by pupils' names or by types of error with references to the folders containing the material itself, one can find what is wanted.

If, in connection with geography, material were secured on the United States Coast Guard Service, this material would be appropriately filed in a folder labeled "Coast Guard Service." The folder thus labeled would be placed in alphabetical position with folders labeled with other topics. It is to be understood that this is the *file* as distinguished from the index. It contains the actual material. Suppose that one piece in this folder was a picture of the launching of the surfboat, and suppose further that this picture has been found to be useful in teaching the geography of the United States. As long as you are interested in gathering the material on Coast Guard Service you do not need an index, but if you are likely to want to collect fifth-grade geography material or pictures, you would have to go through many files in order to do so, unless you had the material index. A picture of the launching of the surfboat might be indexed on a 5″ x 3″ card as follows:

GEOGRAPHY *United States*

Launching
the Surfboat

Coast Guard
Service

On this card "Geography" is the main head and "United States" the secondary head. The entry "Coast Guard Service" directs one to the folder in the file containing material relating to this topic. (Or, if desired, the folder devoted to Coast Guard Service can be given a code number and that number entered on the index card instead of writing "Coast Guard Service.")

In the index drawer—that is, the drawer in which the 5″ x 3″ cards are filed—there should be an equipment of guide cards, at least one for each main heading. A guide card is a card with a raised tab on which a heading may be written. Guide cards are made in various sizes. Those for use with 5″ x 3″ cards, such as I am referring to, are likewise 5″ x 3″ in size. They are usually colored and the commonest colors are blue and salmon. The raised tabs may extend a half, a third, or a fifth of the way across the top of the card. Tabs extending a third of the way across are very generally used. They offer a width of writing surface for the heading amounting to a full inch and a half. In extended indexing, such as I should hope your added professional experience would lead you to adopt, there are both primary and secondary guide cards. They are of different colors, and the secondary guide cards are used to break up a main heading into subordinate headings. Consider, for example, the index card shown on page 96. This card would be filed behind a guide card marked "Geography" and all other cards relating to geography would likewise be placed behind this guide card. If the material were sufficiently extensive to warrant subdivision, there would be a subordinate guide card marked "United States," and the index card to which I am referring would appear among others relating to the geography of the United States behind this subordinate guide card.

Such a filing and index system has among other things its relation to statistics, although it has many other relations as well. If you file your test papers, you will be able to classify them over a period of years under certain headings. You will see how your emphasis in teaching a given subject has shifted from time to time and you can study the question of whether this shift of emphasis was or was not justified. In the study of such a question you are clearly brought face to face with the relation of your emphasis on one topic or another with characteristics of the classes with which you have been dealing. Evidently, then, questions of relationship will come in for consideration. You will want to know how a thirty per cent increase over the usual emphasis on *irrigation* is related to the proficiency of the class. On the one hand, this thirty per cent additional emphasis might be measured by the number of references to irrigation in your planbook and the number of times you asked questions about irrigation in your tests. On the other hand, increased proficiency on the part of the class might be recorded in the greater percentage of correct responses to questions about irrigation, and the number of verbal problems in arithmetic formulated by the pupils on the same topic.

CHILD ACCOUNTING

But there is another class of data which the modern teacher will record and will have at her disposal. Roughly we might divide the records appropriate for the classroom into two kinds: first, those pertaining to instruction, and second, those pertaining to pupils. The first of these I have been discussing in the above paragraphs. The second are no less important.

It is assumed that every teacher keeps a register; in fact there is only one state in the Union which does not require a register to be kept. By "register" I mean a book—or in some instances a set of card forms—in which at least the attendance and punctuality of each pupil is recorded for each day of the term in question. Ordinarily the register also includes for each pupil his date of entry to the class in question, his date of withdrawal from it, his aggregate attendance and total times tardy together with certain scholastic information and a final statement as to his promotion or failure. Quite generally, too, the register includes for each pupil his date of birth (or age upon entering the class), his address, and the name of his parent or guardian.

The register is an exceedingly important school document. It is the court of last resort in legal matters. If a truancy case is contested, it is the teacher's register which settles the fact of the pupil's presence or absence on a given date. Moreover, the register is the basis for all reporting on the part of the principal and the superintendent which involves the attendance, tardiness, late entrance, transfer, causes of withdrawal, scholarship, and promotion of pupils. It may be and very frequently is likewise the original record on ages of pupils, their geographical distribution, and the relation of these to each other and to matters of scholarship, grading, and promotion.

It is evident, therefore, that if the register is not well and truly kept, much of the reporting, by means of which the confidence and support of the public are secured, is vitiated at its source. Accordingly, in some states and in some localities in other states, teachers are not only required to make affidavit to their registers but also to have kept them to the satisfaction of an inspector before they can draw their pay. All such regulations are of course unnecessary if the right attitude on the part of the teachers themselves exists. I think it will exist in most cases when the importance of the register is understood.

You will readily see from the foregoing that in trying to answer the question, "What Statistics Does the Modern Teacher Need?" I include much more than the computations which you may have had in mind when you began to read this chapter. These computations are indeed important, and it is not my purpose to neglect them, but in much of the statistics employed both in education and in other fields there is a regrettable disregard of the basis upon which all statistics must rest—namely, the integrity of the fundamental data. The modern teacher will have respect for and carefully keep the record and report forms which are really worth while in school work. He will consider it worth while to make records of his own and to file the materials and products of his teaching. In preserving these materials he will have a well-developed system of organization.

NUMERICAL DATA

But much of the meaning derivable from specific data, especially when these data become at all extensive, depends not only upon the careful recording and organizing of the data but also upon their numerical treatment. It is to this numerical treatment that I now propose to call your attention.

Broadly speaking, the statistics which the teacher will find useful in his practical work may be divided into three parts —the statistics of *averages*, the statistics of *variation*, and the statistics of *relationship*.

THE STATISTICS OF AVERAGES

The term average has both a general and a specific meaning. In its general sense it means any expression intended to give by means of a single figure the general weight or typical measure of a group of measures. In its restricted sense it is the *arithmetic mean*—that is, it is the sum of the measures divided by the number of them. This is what we generally mean when we speak of *the* average. The arithmetic mean is often called merely the mean. Other averages commonly used in educational statistics are the *median* and the *mode*. Another term for average, in its general sense, is "measure of central tendency."

The mode is the most frequent score or measure found in a series. If, for example, a class takes a spelling test consisting of 25 words and more pupils get 15 than any other number of words right, then 15 is the mode. This average, or measure of central tendency, is not very often used in educational work. With groups of children as small as the ordinary class it is scarcely ever appropriate.

I assume you know what the arithmetic mean is. Most teachers occasionally find the average ratings of pupils. This they do by adding the ratings and dividing by the number of the ratings. This method of obtaining the arithmetic mean will do very well for groups of children such as the teacher is likely to be concerned with.

The median. Of the three averages the median demands special consideration. Its use is very common in educational statistics; but since it has rather recently come into vogue, I shall present it in some fullness.

In the first place, I wish to point out that any average or measure of central tendency must take account of all the items. The arithmetic mean does this admirably. We add the items and then divide by their number. The median, if it is to represent all the items, must likewise take account of them all.

In the second place, a measure of central tendency must, as its name implies, take a middle position. It cannot be extreme. Both the mode and the arithmetic mean tend strongly to take this in-between position. The mode is the measure of most frequent occurrence, and such a measure is usually near the middle of the range of items. Mediocrity is a common condition. In computing the arithmetic mean we, in fact, balance small items over against large ones. Thus we secure a medial value.

But the median as a measure of central tendency is a particularly strong claimant for our favor on this ground of being a mid-measure. In fact, a simple definition of the median is that it is "the middle number when all the numbers are arranged in the order of their size, beginning with either the largest or the smallest."

The median also qualifies as a measure of central tendency in virtue of the contribution to it of all the measures or items. In the orderly arrangement from largest to smallest, or vice versa, *all* the measures are included in the arrangement.

Let us consider the computation of the median. Suppose that in March a certain eighth grade consisting of 38 pupils took a ten-problem test in arithmetic with the following results, the figures indicating in each case the number of problems a pupil got right: 8, absent, 10, 9, 1, 9, 9, 7, 6, 7, 8, 8, 6, 10, 3, 7, 5, 5, 3, 8, 8, 7, 3, 8, 7, 3, 3, 6, 3, 3, 8, 6, 1, 4, 6, 5, 7, 6. Inspection of these scores discloses the fact that the lowest score is 1 and the highest 10. Moreover, we note that 10 problems were gotten right by two pupils, 9 problems by three pupils, 8 problems by seven pupils, etc. A statement of facts such as these may be put in compact form as in Table I.

TABLE I.

No. of Problems Correct in March	No. of Pupils
10	2
9	3
8	7
7	6
6	6
5	3
4	1
3	7
2	0
1	2
Total	37

Such a table is called a frequency table because it shows the frequency with which each measure, or score, in the first column occurred. Each entry in the first column may be called a *step*. Another name for it is a *statistical class*.

Beginning at the bottom of this table, we note that 2 pupils got one problem right; 7, three; 1, four; 3, five; and 6, six. These 2, 7, 1, 3, and 6 pupils make 19. Since one pupil was absent from this test there were 37 present. The middle score when the scores are thus arranged in order would therefore be the nineteenth. The nineteenth score turns out to be 6; and on this basis—that is, on the basis of the median as the middle number when all the numbers are arranged in the order of their size—6 is the median.

But the method I have just given you for obtaining the median is a rather rough method. Moreover, it is not the one which you will ordinarily be expected to use when you give standardized tests or other examinations to your pupils. I have, however, given you this simplified method not only because it is sometimes used but also because it serves to bring to your notice the fundamental principle on which the median is based, namely, that it represents a middle position among the items. A more refined definition of the median considers it as a *point* on the range of the measures on either side of which there is the same number of measures. The method by which the median, thus considered as the mid-point, is obtained may be called the mid-point method.

In June the same eighth-grade class to which I referred above took another problem test in arithmetic with the following results, each figure indicating the number of correct problems: 8, 9, 9, 10, 9, 9, 9, 10, 10, 10, 6, 10, 9, 9, 10, 9, 10, absent, 9, 9, 6, 10, 3, 10, 7, 7, 5, 8, 7, absent, 8, 8, 1, 5, 6, 8, 7, 8. Arranging these numbers as before we have the frequency table shown as Table II.

TABLE II.

No. of Problems Correct in June	No. of Pupils
10	9
9	10
8	6
7	4
6	3
5	2
4	0
3	1
2	0
1	1
Total	36

You will notice that 36 pupils are represented in Table II. Therefore, in seeking to obtain the median by the mid-point method we shall naturally try to locate a point such that the 18 highest scores may be on one side of it and the 18 lowest scores on the other side of it. A rather obvious procedure will be to start at the lowest score and count until we reach the eighteenth score. One pupil got one problem right, one got three right, and two got five right. So far this accounts for the four lowest scores. We are still, however, far from the eighteenth score. Adding in the three scores at 6, the four at 7, and the six at 8 will give us the 17 lowest scores. We need one more to make the required 18. Clearly this will have to be one of the 10 scores at step 9, and it must be, theoretically at least, the lowest one of these 10 scores.

Here we make an assumption: namely, that these 10 scores at 9 are distributed evenly from *just* 9 up to but not including 10. This assumption is based on the idea that among the 10 children who got 9 problems right, there were probably differences in ability all the way from that of the child who was just able to do 9 problems to that of the child who fell just short of getting 10 problems. In between these two extremes there are other children whose abilities would be represented by scores between 9 and 10. The mid-point method of obtaining the median not only recognizes fractional measures of ability but also assumes that the abilities are evenly spaced within the limits of the step in question—in this case within the limits of the score 9, which as already explained is from just 9 to just short of 10.

Accordingly, we shall find that of the 10 scores assumed to be distributed between 9 and 10, the first one, or the eighteenth score from the low end of the entire range, is 9.1. This, therefore, is the median.

Let us return now to the table showing the success of these eighth-grade pupils in the March test. The table is repeated on the opposite page.

On this occasion 37 pupils took part. Half the number of measures, therefore, is neither 18 nor 19. Here we adopt a convention in virtue of which half-measures are recognized. In seeking the point above and below which there is the same number of measures, we shall therefore attempt to locate the point arrived at by counting up to measure No. 18½. Counting from the bottom up we find that 13 pupils or scores carry us through the three pupils who got

TABLE I (Repeated).

No. of Problems Correct in March	No. of Pupils
10	2
9	3
8	7
7	6
6	6
5	3
4	1
3	7
2	0
1	2
Total	37

five problems right. In order to count up to 18½ we shall need 5½ more measures. These we must obtain from the next 6 measures. Since these are located at step 6, this step is called the median step. If these are assumed to be evenly spread over the interval between just 6 and just less than 7, the 5½ lowest of them will carry us 5½-sixths of the way through this interval. Our computation is then completed when we have added to 6 the complex fraction 5½-sixths. Now 5½-sixths when changed to a decimal may be expressed as 0.9. Adding this to 6 gives 6.9, which is the median.

To sum up, then, the procedure in computing the median by the mid-point method is as follows: first, arrange the measures in a table (called a frequency table) in the order of their size. Second, find the total number of measures, in this case, the total number of pupils. Third, divide this total number by two; the quotient may be called the *half-*

sum. Fourth, total the frequencies from the low end until the inclusion of the frequencies at the next step (which is the median step) will exceed the half-sum already obtained. (In the case of the June scores, this carries us through the six frequencies at score 8, and in the case of the March scores, it carries us through the three frequencies at score 5.) Fifth, subtract the total frequencies to this point from the half-sum, thus obtaining the remaining frequencies to be considered. Sixth, form a fraction whose numerator is this remainder and whose denominator is the number of frequencies at the median step. Seventh, add this fraction to the lower limit of the median step; the result is the median.

Problems.—Find the median from the following frequency table, shown as Table III.

Applying the seven operations listed in the above summary, we find that the first and second operations have been completed. From that point the solution proceeds as follows:

(3) $124 \div 2 = 62$.

(4) If we total the 7 lowest frequencies (2, 3, 5, 3, 25, 7, and 8) we get 53. If we were to add the 13 at step 20 we should exceed the half-sum which we have already obtained, namely, 62. Accordingly, 53 is the total of the frequencies called for in this operation, and step 20 is the median step.

(5) $62 - 53 = 9$.

(6) Form the fraction 9/13. This may be done as I have indicated in the table itself. This means, of course, that to arrive at the median we must proceed 9/13 of the way from 20 toward 21, in other words, 9/13 of the way through the step 20.

(7) Add 9/13 (changed to a decimal) to 20, obtaining 20.7 as the median.

TABLE III.

ACHIEVEMENT OF FIFTH-GRADE PUPILS IN SPELLING WORDS FROM COLUMN L OF AYRES' SCALE FOR MEASURING SPELLING ABILITY

No. of Words Spelled Correctly	*f* (Frequency)
25	7
24	5
23	11
22	14
21	21
20	9/13
19	8
18	7
17	25
16	3
15	5
14	3
13	2
Total	124

You will observe that in the foregoing tables the steps have differed from each other by 1. If, however, the test had been a spelling test of 50 words and if we had shown in the first column every possible score, the table would probably have strung out to an inordinate degree. A certain class, for example, yielded the following spelling scores as the result of a 50-word test, the figures indicating the number of words spelled correctly:

46, 18, 30, 28, 28, 25, 20, 38, 42, 39, 38, 37, 36, 37, 35, 35, 31, 30, 29, 27, 26, 22, 21, 21, 20, 35, 29, 27, 13, 13, 8, 42, 41, 24, 6, 5, 18, 38, 22, 27, 21, 21, 1, 32, 5, 9, 14, 9.

In this layout 46 is the highest score and 1 is the lowest. If you make a frequency table in which you show tne number of times each score between 46 and 1 occurred, you will have a very long and very unsatisfactory table. But if you group the scores by 5's and show the frequency of scores within each group of 5, you will obtain a much better table. Table IV shows how it will appear.

TABLE IV.

No. of Words Correct	No. of Pupils
45–49	1
40–44	3
35–39	10
30–34	4
25–29	3/9
20–24	9
15–19	2
10–14	3
5–9	6
0–4	1
Total	48

In such a case the procedure in finding the median is exactly the same as I have already indicated to and including the sixth operation. In this instance after you have formed the fraction called for by that operation, you obtain the fraction 3/9 (I have indicated this in the table.) This fraction is then multiplied by 5 because the step is 5 wide. This gives 1 2/3 or, expressed decimally, 1.67. This amount is added to the lower limit (25) of the median step, giving 26.67 as the median. Don't forget to multiply the fraction

PROBLEM 1.

No. of Addition Examples Attempted	No. of Pupils (Frequency)
12	1
11	0
10	1
9	2
8	9
7	2/10
6	9
5	5
4	4
3	1
Total	42

Median = 7.2 examples attempted.

PROBLEM 2.

No. of Examples Correct, Courtis Practice Tests	No. of Pupils (Frequency)
18	1
17	0
16	0
15	3
14	0
13	4
12	6
11	1.5/7
10	10
9	4
8	3
7	0
6	0
5	1
Total	39

Median = 11.2 examples correct.

by the number of units in the median step when that step has more than one unit in it.

For practice I give here and on the preceding and following pages several frequency tables which you may use in computing the median. In each case I have given the correct answer.

PROBLEM 3.

Words Spelled	Frequency
48–49	3
46–47	5
44–45	7
42–43	10
40–41	5/12
38–39	8
36–37	7
34–35	6
32–33	5
30–31	1

Median = 40.8 words.

PROBLEM 4.

Marks in Latin	Frequency
95–100	5
90–94	8
85–89	12
80–84	14
75–79	22
70–74	14
65–69	8
60–64	7
55–59	5
50–54	3
45–49	2

Median = 77.5.

PROBLEM 5.

Monthly Salaries Paid to Teachers of History.

Salary	Frequency
$145.00–$154.99	1
135.00– 144.99	1
125.00– 134.99	2
115.00– 124.99	2
105.00– 114.99	3
95.00– 104.99	10
85.00– 94.99	15
75.00– 84.99	14
65.00– 74.99	5
55.00– 64.99	1
Total	54

Median = $89.67.

The median possesses certain advantages which the arithmetic mean—or the average, as commonly designated—does not possess. One of these advantages, and it is a very important one in our work, is that the median may be applied in cases where the items are merely ranked. The ranking may even be faulty in some respects and a median may still be derivable. The advantage of this availability of the median may be illustrated in the case of compositions. Even with the best composition scales it is by no means easy to rate the written work of children. If we desire to obtain the average performance of a class in respect to a given set of compositions, we may avail ourselves of our undoubted ability to distinguish between two compositions as to their merit. We know that we can do this without measuring them. The fact that we can enables us—at least in theory—to arrange a given set of compositions in the order of their excellence. When this has been done, the middle composition is the median composition, or, if there are an even number of papers, either one of the *two* middle compositions may be taken as the median performance. The difference between two adjacent compositions in the middle of the package will be so small that no material error will be made in taking either of them for this purpose. Nevertheless it would probably be wise to have a consistent practice in this regard and I suggest that the theoretically better one of the middle two compositions (when there are an even number of them) be taken as the median performance.

The median may be applied in this manner—that is, by arranging performances in the order of their merit—to a number of different kinds of school work. It may be applied to handwriting, drawing, manual training, sewing, cooking—in fact to any subject where the product of each pupil can be ranked according to merit. If the manual-training class makes tabourets, the tabourets may be arranged from the best to the poorest, and the middle one may be exhibited and preserved as the typical or average performance of the class.

In the case of composition, certain interesting possibilities grow out of this use of the median. Having arranged the compositions for the first week of the term in the order of their merit, and having selected the median composition, we may place it on a chart or a bulletin board as the *average achievement of the class*. We may even rate it by means of one of the composition scales; and a group of teachers might coöperate, each to rate the others' median compositions for that week, thus permitting an average of several ratings to be taken with reference to each median composition. The next week the same procedure may be followed, and the median composition may be put on the chart alongside of the first one. If this is continued throughout the semester, there will be a picture of the progress of the class during the term in composition writing—this picture being composed of successive typical performances of the class.

The arithmetic mean. I wish to show you very briefly the so-called short method of computing the mean. I have already indicated that the arithmetic mean is the sum of the measures divided by their number. This is the ordinary way of computing it. There is, however, another property of the mean to which I wish to draw attention. Since, as has already been indicated, the mean takes a middle position, some of the measures are greater than the mean and some are less than it. Call the amount by which a measure exceeds the mean its *plus deviation,* and in the case of a measure less than the mean call the amount by which it is less its *minus deviation.* Now the property of the mean to which I am referring is that the sum of the plus deviations from it equals the sum of the minus deviations from it. This property is used in the short method of computing the mean. It may be illustrated concretely as follows. Suppose that in bowling I make three scores—namely, 150, 156, and 159. What is my mean or average? Of course one way of obtaining this is to add the numbers and divide by three. Another way would be to *guess* an average and then to correct it so as to make the plus and minus deviations the same. Suppose I guess that the average is 154. Then the minus deviation of 150 is 4. The plus deviations of 156 and 159 are 2 and 5 respectively, giving a total of 7. Thus the plus deviations exceed the minus deviation by 3, and the guessed average is, therefore, too low—too low because it makes the minus deviation too small and the plus deviation too large. The amount by which it will have to be raised is exactly indicated by the discrepancy between the plus and minus deviations (namely, 3) when this is divided by the number of items. In other words, the discrepancy of 3 is distributed among three items and therefore indicates that the average must be raised one unit—namely, from 154 to 155. The mean is therefore 155 and this is the same as would be secured by adding 150, 156, and 159 and dividing by 3.

Consider now the scores already given in Table I on page 103 of the 37 eighth-grade children who took an arithmetic problem test in March. These are given in Table V. The first two columns of Table V reproduce these scores and the frequency with which each one of them was obtained.

Let us guess an average and find the deviations of all the measures from it. Let us assume that the average number of problems correctly solved is 7. Now, on the assumption that 7 means from just 7 to just less than 8 the most probable distribution of ability among the pupils having 7 problems right is an even distribution from 7 toward 8. The average of these abilities will be 7.5 and therefore when we guess our average to be at step 7 we must represent it as 7.5. Similarly the ability represented by 6 is regarded as 6.5, that represented by 5 is regarded as 5.5,

TABLE V.

Scores	No. of Pupils (*f*)	Deviations (*d*)	(*f*) × (*d*) or (*fd*)
(1)	(2)	(3)	(4)
10	2	+ 3	+ 6
9	3	+ 2	+ 6
8	7	+ 1	+ 7 + 19
7	6	0	0
6	6	— 1	— 6
5	3	— 2	— 6
4	1	— 3	— 3
3	7	— 4	— 28
2	0	— 5	0
1	2	— 6	— 12 — 55
Total	37		— 36

$$-36 \div 37 = -0.97 \ (c).$$

etc. In other words, the frequencies in each step of the distribution are regarded, for the purpose of finding the mean, as concentrated at its mid-point.

And now for the deviations. The 6 pupils at 7 (or, according to our assumption, 7.5) clearly do not deviate at all from the guessed average of 7.5. Accordingly we may place a zero in the third column of Table V opposite 6. This column is headed "deviations," the abbreviation for which is *d*. The 7 pupils who got 8 problems right (regarded as 8.5) deviate in each case by plus 1 from the guessed average of 7.5. The three pupils at 9 each deviate from the guessed average by plus 2, and the two pupils at 10 deviate by plus 3. The 6 pupils who got 6 problems right deviate from 7.5 by minus 1, those who got 5 problems right by minus 2, etc. These deviations you will find in the third column of Table V.

In order to find out the amount of plus deviation and the amount of minus deviation which this column exhibits, we must multiply the entries in Column *3* by those in Column *2*, giving a column which we may call the *fd* column. Two pupils *each* deviate by +3, which makes a total deviation for them of +6. Similarly the other products are entered in the *fd* column of Table V. The sum of the plus deviations is now found to be 19, and the sum of the minus deviations is found to be 55. Evidently the guess of 7.5 was a poor one. Is it too small or too large? At this point it will be well for you to do a little thinking. The minus deviations are considerably too many. Therefore, in order to reduce them, the true mean must clearly be closer to these small numbers. In other words, it must be smaller than we have guessed it to be. The amount by which it must be smaller is now easily obtained. The algebraic sum of Column *4* is minus 36. As in the case of the bowling scores this discrepancy between plus and minus deviations is to be distributed to all the measures by dividing it by the number of measures. When —36 is divided by 37 the quotient is —0.97. This is the amount by which the guessed average of 7.5 must be reduced in order to obtain the true average. The true average is, therefore, correct to the second decimal place, 6.53. Since the quotient —0.97 is used to *correct* the guessed average, it is called the correction, which may be abbreviated to *c*.

If the algebraic sum of Column *4* had been plus instead of minus, in other words, if the correction had been positive, it would have been *added* to the guessed average to obtain the true average. We may sum up by saying that if the correction is minus, you subtract it from the guessed average and if it is plus, you add it to the guessed average.

Consider the following problem. What is the mean intelligence quotient of the 107 children represented in Table VI? Note that the steps in Column *1* are 10 wide, that is, "130-139" means from just 130 to just short of 140, and similarly for the other steps. The most promising place for the mean seems to be either at the step "110-119" or the step "100-109." Let us choose the latter. In reality the final result will be the same no matter what step we select for the guessed average. Remember that in this method all the measures at a given step are supposed to be concentrated at the mid-point. Therefore, guessing the mean at "100-109" means guessing it at the mid-point of this step, or at 105. The deviations are as indicated in Column *3* of Table VI and the products are as shown in the fourth column. The sum of the positive deviations is +68 and that of the minus deviations is —48. Accordingly the algebraic sum is +20. +20 divided by 107 gives +0.187. This is the correction, but it is in terms of steps because the deviations as entered in Column *3* are in steps. But we have guessed the average not in steps but in units of I. Q. Consequently before we can use the correction we must convert it into units of I. Q. Since each step has 10 of these units, we need simply multiply +0.187 by 10. This gives +1.87. Whenever you are working with data that are grouped by steps larger than 1, do not forget to multiply the correction by the size of the step. *This is the place where you will be most likely to make a mistake.*

Since the correction is plus we add it to the guessed average, which is 105. Adding 1.87 to this gives 106.87 as the average or mean.

TABLE VI.

I. Q.	No. of Pupils (*f*)	Deviations (*d*)	*fd*
(*1*)	(*2*)	(*3*)	(*4*)
130–139	5	+ 3	+ 15
120–129	15	+ 2	+ 30
110–119	23	+ 1	+ 23 + 68
100–109	25	0	
90–99	31	− 1	− 31
80–89	7	− 2	− 14
70–79	1	− 3	− 3 − 48
Totals	107		+ 20

$$+20 \div 107 = +0.187 = c \text{ (in steps).}$$
$$+0.187 \times 10 = +1.87 = c \text{ (in units of I. Q.).}$$
$$105 + 1.87 = 106.87.$$

I suggest that you now turn back to the problems on pages 111-113. Compute the means for each of these problems by the short method. The following are the answers to these problems, correct to two decimal places:

Problem 1. 7.17 examples attempted.

Problem 2. 11.50 examples attempted.

Note: Observe that if the guessed average is at 11.5, the correction is zero. In other words the guessed average is then the true average.

Problem 3. 40.56 words.

Note: Observe that the steps are two units wide. If you guess the average as being in the step "40-41" you are guessing that it is 41, because 41 is the mid-point of this step (running from 40 through 41 up to but not including 42).

Problem 4. 76.70.

Note: Here each step is five units wide. If your guessed average is at the step "75-79" it is at 77.5 because 77.5 is the mid-point of this step.

Problem 5. $92.59.

Note: The steps here are ten units wide. The mid-point of step "95-104.99" is 100; the mid-point of step "85-94.99" is 90.

I have already pointed out that the field of elementary statistics may be roughly divided into the statistics of averages, the statistics of variation, and the statistics of relationship. The statistics of averages is especially concerned with the three measures of central tendency to which I have been calling your attention, namely, the arithmetic mean, the median, and the mode. You will have far less use for the statistics of variation and still less for the statistics of relationship. The ideas of relationship and of variation are common enough, but at present school practice scarcely takes account of them in a statistical way. I shall, therefore, content myself with a very brief statement as to the common measures of variation and relationship.

STATISTICS OF VARIATION

While it may not often be necessary for you to compute a measure of variation, the idea of variation is really fundamental. When you have obtained an average you have not in reality secured a highly satisfactory figure. It is true that the average is frequently employed and has the sanction of usage and of general familiarity. But two series of measures—two frequency tables such as those which we have been considering—may have the same average, yet be quite different in their make-up. For example, the scores of the pupils in one class may cluster closely around the average of the scores, and in that case the average represents all the pupils very well. The scores of the pupils in another class may yield the same average, yet these scores may be widely scattered, running from very low to very high. Under these circumstances the average will not serve nearly so well as in the first case to typify the performance of all the pupils.

And does not the difference between these two classes, whose averages are the same, constitute a very real difference? Is not the teaching of pupils who are all pretty much of the same ability very much easier than the teaching of pupils who differ widely from each other in ability? Is the difference in average ability between two classes as important, from a teaching point of view, as the difference between classes in point of variation in ability? Would you not rather teach pupils of about the same level, even if that level were low, than pupils whose higher average ability was accompanied by a range from very bright to very stupid?

It is undoubtedly true that the use of measures of variation would be greater among teachers if they knew more about them and were able to compute them. You would do well to secure a book on statistics and study the chapter on variation.

Three measures of variation are in general use. These are the quartile deviation, the standard deviation, and the mean deviation. The *range* has also been suggested as a measure of variation, but although it is easily understood it has little else to recommend it. If pupils take a spelling test and the lowest score is 5 while the highest is 40, the range is 35.

Quartile deviation. When we arrange the measures in the order of their size—as for example in Problem 4, page 112—we can count in from the lowest score one-quarter of the way with no more trouble than we experienced in counting half of the way for the purpose of getting the median. Indeed, we may count any fractional part of the way we please. But one-fourth of the way has a certain importance, as I am about to show; and so also has three-fourths of the way. Now, just as the point reached when we count in half-way is given a special name, the median, so the point reached when we count in the first quarter of the way has a name. It is called the *first quartile.* Similarly the point reached when we count in three-fourths of the way is called the *third quartile.* The median is really the second quartile, although it is never so called.

In Problem 4 the total number of pupils (namely, the total frequencies) is 100. One-quarter of 100 is 25. If we count in from the bottom until we have 25 frequencies, we pass completely through the step 65-69—that is, we arrive *at* 70. The first quartile, therefore, is 70. Continuing our count three-quarters of the way, that is, until we have totaled 75 frequencies, we shall find that we have gone completely through the step 80-84, arriving at 85. The third quartile, then, is 85. The quartile deviation is defined as *half the difference between the third quartile and the first quartile.* In this case, therefore, it is half the difference between 85 and 70, or 7.5. The quartile deviation may be thought of as half the range within which 50 per cent of the measures or scores lie. In Problem 4, 50 per cent of the scores lie between 70 and 85; and half the range between 70 and 85 is 7.5, the quartile deviation.

Standard deviation. The standard deviation is theoretically the most satisfactory of all the measures of variation. It measures the amount of dispersion about the mean. In other words, one finds the arithmetic mean, then finds the amount by which each measure (score) differs from this mean, and then finds a sort of average for all these differences or *deviations*. I say "a sort of average," for it is not one of the kinds of averages which we have been considering. We first square the deviations, then sum these squares, and then extract the square root of the sum. The reason for this is largely a theoretical one having to do with the mathematics of statistics. One may illustrate the meaning of the standard deviation by a very simple problem. Suppose that five pupils answered correctly the following numbers of questions in a geography test: 11, 15, 18, 6, and 10. The arithmetic mean of these scores, by the ordinary process, is 12. The deviations of these five pupils from 12 are respectively: -1, $+3$, $+6$, -6, and -2. To obtain the standard deviation, square each one of these deviations. This will give 1, 9, 36, 36, and 4. Find the sum of these squares. It is 86. Extract the square root of 86. It is 9.3 to the nearest first decimal place. This is the standard deviation. It would not ordinarily be computed for so small a number of measures as five. My illustration is merely an illustration.

Mean deviation. The mean deviation is also a very satisfactory measure of variation. Like the standard deviation, it shows the dispersion of the measures *around* or *from* the measure of central tendency. Thus, as in the case of the standard deviation, one first finds an average, but this time it is better procedure to use the median than to use the mean. All one does after finding the median is to find the difference between each measure and the median, add these differences, disregarding their sign, and divide by the number of measures. Consider, for example, a class of 12 pupils who have obtained the scores indicated in Table VII in a Courtis Arithmetic Test in addition.

TABLE VII.

Examples Right	Frequency
12	1
11	3
10	4
9	2
8	2
Total	12

The median is 10.5. According to the mid-point method, the ability of the one pupil who scored 12 is to be represented by 12.5. Accordingly his deviation from the median is 2. Similarly the three pupils who got 11 examples right and whose ability is represented as 11.5 *each* deviate from the mean by +1. Since there are three of them this means a deviation of +3. The two pupils at 9 (regarded as 9.5) each deviate by —1, and the two pupils who got 8 problems right each deviate by —2. Summing all these deviations without regard to sign, we find that we have 11. This, divided by 12, the number of cases, gives 0.9 to the nearest first decimal place. The process will be easy to understand from Table VIII, which is an extension of Table VII.

TABLE VIII.

Examples Right	Frequency	Deviation	Frequency Times Deviation
12	1	+ 2	+ 2
11	3	+ 1	+ 3
10	4	0	0
9	2	− 1	− 2
8	2	− 2	− 4
Total	12		11

11 ÷ 12 = 0.9, mean deviation.

STATISTICS OF RELATIONSHIP

In the statistics of relationship we attempt to express the extent to which an increase in one ability—or whatever it is we are measuring—is accompanied by an increase or decrease in another ability—or thing measured. If, for example, ability in geography "goes with" ability in history, the extent to which it does so will be a measure of relationship; and in the process of securing this measure we are using the statistics of relationship.

There is a rather large number of ways in which relationship may be measured. For example, we may make use of the idea of rank. According to this idea, perfect direct relationship will exist when a pupil who is first in history, let us say, is also first in geography; when the pupil who is second in history is also second in geography, and so on, each pupil being of the same rank in the two subjects.

But if the pupil who is first in history is no more likely to be first in geography than to be of any other ranking, and if the same lack of correspondence is shown for the pupil who is second in history, the pupil who is third, and so on, then we have no relationship. Finally, if the pupil who is first in history is last in geography, and the pupil who is second in history is next to the last in geography, and so on, with complete inversion of ranks, then the relationship is just as close as it was in the first case, only instead of being direct it is inverse.

We have a way of measuring relationship according to rank. It will do as an approximation and may be used when not many pupils are involved. Consider, for example, a class of 12 boys, each of whom obtained scores in geography and history as given in Table IX.

TABLE IX.

Pupil	Geography Score	History Score
A. B.	10	4
C. D.	16	8
E. F.	17	9
G. H.	13	8
I. J.	15	11
K. L.	12	9
M. N.	12	8
O. P.	11	7
Q. R.	16	7
S. T.	13	5
U. V.	8	1
W. X.	18	12

Now suppose we rank these pupils. The result will be as indicated in Table X, page 128.

Table X.

Pupil	Geography Rank	History Rank
A. B.	11	11
C. D.	3.5	6
E. F.	2	3.5
G. H.	6.5	6
I. J.	5	2
K. L.	8.5	3.5
M. N.	8.5	6
O. P.	10	8.5
Q. R.	3.5	8.5
S. T.	6.5	10
U. V.	12	12
W. X.	1	1

Perhaps a word of explanation about the ranking may not be amiss. You may not infrequently have occasion to rank your pupils for one purpose or another. Two or more pupils often have the same score, and the question of how to rank these pupils must be decided. For example, in the above table W. X., whose score is 18, is clearly rank 1. E. F., whose score is 17, is equally clearly rank 2. But C. D. and Q. R. each obtained 16. They are equally entitled to rank 3 and rank 4. Since they share these two ranks, and since neither pupil has any stronger claim than the other, they are each given a rank midway between the two ranks which they share—namely, 3.5. They have together occupied ranks 3 and 4. Consequently I. J., who has the next lower score, has rank 5. Then come G. H. and S. T., each with score 13. They share ranks 6 and 7; consequently each is given the rank 6.5. In history, after the first four ranks have been assigned, it is found that three pupils each have the next highest score. These are C. D., G. H., and M. N., each of whom scores 8. Since they share equally in ranks 5, 6, and 7, they should all be given as a rank the average of 5, 6, and 7, or 6.

If perfect correlation is indicated—so far as a rank method can indicate it—by exact correspondence of rank, pupil A. B. with rank 11 both in geography and history evidently contributes by his performance to close and direct correlation. But C. D. does not contribute so much. He ranks 3.5 in geography and 6 in history. The discrepancy is evidently 2.5, and it seems pretty clear that a measure of the amount of relationship exhibited by these 12 pairs of scores is going to have something to do with discrepancy like this—with differences, in other words, between the ranks in geography and history. In the case of C. D. the number indicating the rank in history is larger than the one indicating the rank in geography. Sometimes, as in the case of K. L., the larger figure is in geography. If we call one sort of difference a plus difference, the other must be a minus difference. We may, however, disregard all the differences except those in one direction. In other words, we may give attention only to the differences where the number indicating rank in history is *larger* than the number indicating rank in geography. If now we find these differences they will run as follows:

C. D.	2.5
E. F.	1.5
Q. R.	5.0
S. T.	3.5

If these one-way differences in rank are added, we find that their sum is 12.5.[6] Evidently this tells us something about the amount of relationship. We know that if the relationship were perfect and direct, the sum of the differences would be zero, because there would not be any differences in rank. Now, if we knew what the sum of the differences in rank would be for entire absence of correlation, we could express the per cent that our 12.5 is of that sum of differences. It is provable that if these pupils exhibited no relationship whatever between geography and history, the sum of their differences in rank taken one way would be $1/6(n^2-1)$, where n is the number of pupils, in this case 12. This means that absence of relationship would be indicated by 23.8 ($144-1=143$; 1/6 of $143=23.8$). 12.5 is 53 per cent of 23.8. In other words, the lack of relationship is 53 per cent of utter lack of relationship. This may be converted into a positive statement by taking the complement of 53 per cent, namely 47 per cent. If there is 53 per cent lack of relationship, there must be 47 per cent of relationship.

Another name for relationship is *correlation*, and the per cent of relationship—or, as more usually expressed, the decimal fraction of relationship—is called the *correlation coefficient*. The formula for the correlation coefficient by the method I have just described (called the *foot-rule method*) is:

$$R=1-\frac{\text{sum of one-way differences}}{1/6\ (n^2-1)}$$

---

[6] If we had made the subtractions the other way, that is, of geography *on* history, they would have been 0.5, 3, 5, 2.5, and 1.5—total 12.5 as before. It will always be found that the sum of the differences one way is equal to the sum of the differences the other way.

There are other bases of computing correlation. One expresses the ratio between each pair of measures—after making certain adjustments—and then finds the median of these ratios. Another attaches to each measure a sign, plus or minus, according as it is above or below the median of the series of measures to which it belongs. Correlation is then worked out on the basis of the per cent of unlike signs in the pairs of measures. In the problem given above, if a child is above the median in geography his score is plus. If he is below the median in history, his score is minus. Q. R. is such a pupil; and his scores present unlike signs. If there are many such instances there is decided lack of direct relationship.

I have tried in a very simple way to illustrate the meaning of correlation. Each of the methods I have suggested is inferior to the so-called product-moment method, the essential superiority of which is that it takes full account of both the amount and the direction (plus or minus) of the deviation from the measure of central tendency. This product-moment method is one which you should look up in a textbook on statistics if you ever have occasion to use it. It would add too much to the extent of this chapter if I were to explain it in sufficient detail to make its computation and meaning understandable.

I am adding the names of four books on statistics as applied to education:

Odell, C. W. *Educational Statistics.* New York: The Century Company, 1925. 330 pp.

Otis, A. S. *Statistical Method in Educational Measurement.* Yonkers, New York: World Book Company, 1925. 337 pp.

Rugg, H. O. *Statistical Methods Applied to Education.* Boston: Houghton Mifflin Company, 1917. 410 pp.

Thurstone, L. L. *Fundamentals of Statistics.* New York: Macmillan Company, 1925. 237 pp.

CHAPTER III

BETTER TESTING: INTELLIGENCE TESTS

When a person engages in a certain line of work and becomes at all interested in it, he at once seeks a way to estimate his progress and to predict his success. This seems to be a fundamental human need. We naturally desire to see how much and how well we have done. Moreover, only when we have some sense of success do we appear to be able to put forth sustained and vigorous effort.

What may be said of individuals may likewise be said of institutions. The need for measuring the success of different procedures and for comparing the product of different groups of people is everywhere recognized. It is so with the school; it has always been so. Very likely teachers have from the beginning tried to measure their work. Not only would they want to know the status of their pupils, but they would also want to convey this information to the pupils themselves and to their parents.

But something was lacking. In the first place, qualitative terms alone were unsatisfactory. To say that a child was excellent in geography or fair in arithmetic had little meaning unless these terms could be given a true numerical value. In the second place, the units which would make either quantitative or qualitative statements meaningful did not exist. No one knew what a unit ability in handwriting might be nor what it meant for one pupil to read twice as well as another. You have, of course, used the system of percentages or percentage grades by which it has long been pretended that the attainments of pupils could be measured. To the unthinking these percentages seemed to afford a precision of statement which was in fact wholly lacking in them.

THE UNRELIABILITY OF PERCENTAGE JUDGMENTS

Among different judges. I shall not repeat at length the story which has been so often told of the inability of teachers—or of anyone else for that matter—to rate pupils in terms of percentage grades. Their disagreement in judging the value of a given examination paper has frequently been pointed out. For example, a certain county superintendent distributed copies of a pupil's answer paper in an arithmetic examination to one hundred of his teachers. These teachers were asked to rate the paper on the usual percentage basis. One teacher thought it was worth only 10. At the other extreme were four teachers who thought it was worth 100. The remaining 95 teachers rated it between 10 and 100; thus the various estimates ranked from 10 to 100 per cent. If 75 had been the passing mark, half the teachers would have failed the pupil and the other half would have passed him.

Educational writings afford many such accounts—accounts of how teachers have failed to agree even approximately on the value of examination papers. Brooks in his *Improving Schools by Standardized Tests* [1] tells how a group of twenty-four teachers rated the same answers to ten problems in percentage. They gave marks ranging from 65 to 94. Lively argument ensued and the teachers asked for a chance to try again. They were then given an eighth-grade history paper on which they worked hard for half an hour. Curiously enough, the highest and lowest ratings were the same as in the case of the arithmetic paper.

---

[1] Boston: Houghton Mifflin Company, 1923. pp. 17-18.

Do not suppose that these are exceptional cases. If you and some of your colleagues desire to convince yourselves of the need for better testing, you can do so by rating a few papers and observing the extent to which you disagree.

The same judge at different times. In rating a given piece of work, not only do people disagree with each other, but the same person disagrees with himself at different times. Ashbaugh submitted a pupil's answer paper in arithmetic to forty-nine students in education, many of them being teachers of experience.[2] Each student rated the paper three times at intervals of four weeks. Only one of the forty-nine students gave the paper the same score on all three occasions. Only eleven of them succeeded in agreeing with themselves two out of three times. One student scored the paper 55 on the first occasion, gave it 30 on the second occasion, and came back to an in-between position on the third occasion by rating it 40. Another student gave the paper a grade of 54 at the first trial, following it at the second trial with a grade of 76, and concluding at the third trial with a grade of 70.

Observe that these two students agreed almost perfectly in their first rating of the paper. One of them rated it 55 and the other 54. If we were to confine our attention to the first trial—and one rating is all we ever get in the schools—we should say that here were two people who agreed remarkably well in rating this paper. The second trial, however, tells a different story. One person lowers his grade from 55 to 30; the other raises his grade from 54 to 76. The result is that they are now 46 percentage points apart. If the second rating happened to be the only one at our disposal, we should say that these same two persons differed widely in their estimate of this paper. Even on the third trial when each student had brought his rating into closer agreement with his first estimate, the difference was still thirty percentage points.

---

[2] Ashbaugh, E. J. "Reducing the Variability of Teachers' Marks," *Journal of Educational Research*, 9:185-98 (March, 1924).

Without laboring with the point any further, we may, I think, agree that the unassisted judgments of teachers as to the value of a particular piece of work are highly unreliable —at least when judgments are expressed in per cents. Teachers differ among themselves, and each one differs with himself under different circumstances and at different times.

THE IMPORTANCE OF PERCENTAGE GRADES

Yet what tremendous importance do we attach to these grades! They determine the progress of pupils in school. Without attaining to the highly variable and elusive standard of some teacher the pupil may not be promoted. He must meet the same conditions in the high school and college. He must earn his points or credits, if he earns them at all, according to the judgment of one teacher after another, no two of whom could be expected to agree as to the value of a given sample of his work. These points and credits are the things which determine his success as a student, and since studentship is the chief business of his childhood and adolescence, they determine his success as a boy and as a young man. If his grades are satisfactory and he has enough of them, he may enter upon adulthood with certain prospects of success. If he has failed to obtain these grades, he takes his place in adult society with a handicap.

The truth is that if the parents of our pupils really understood the nature of the grades we give in school, they would compel us to develop a greater expertness in estimating products, processes, and materials and would necessitate among other things the adoption of "better testing." A caustic critic has likened the teacher to a Roman augur. The augurs of ancient Rome observed the flight of birds and were supposed to derive from their flight omens with respect to important undertakings. The procedure of the augur was highly technical, and a special vocabulary was devoted to it. At the conclusion of his observations the augur would solemnly declare, "The birds allow it," or he would announce that the proposed action must take place "on another day." The writer who has likened teachers in their rating of pupils to augurs in their observation of the flight of birds points out that the augurs must have known that their practice was a swindle, for all these men were of high birth and attainment. How one Roman augur, he says, could meet another in the street without a grin or a wink or a thrusting of the tongue in the cheek was a mystery. How one teacher, he continues, can seriously talk to another about grades without similar indications is likewise a mystery.

All this may be true, and yet when a pupil is rated over a series of years in several subjects and by a considerable number of teachers, the net result of the estimates may be substantially correct. Teachers may indeed vary with one another and with themselves in rating a particular performance or product, such as an examination paper, without differing greatly in their general estimate of a student's work. There is reason to believe that a certain rough justice is done in the long run. This, however, is far from satisfactory. Even if it were, the "long run" is often impossible. Students frequently remain under a given teacher only a short time. Moreover, justice secured by a series of injustices may have little value. The English teacher may rate too high and the Latin teacher too low. Both are therefore wrong, and the fact that a true average for both subjects may be calculated does not make either of them right. A just estimate of a pupil's attainment in a course in American history, if it is arrived at by underrating a portion of his work and overrating another portion, is clearly inferior to a just estimate which has been attained through a series of accurate ratings.

Enough has probably been said to show the need for better testing. Moreover, when better testing is available, a new opportunity is presented to the teacher. He may become a student of his calling in a sense which was formerly impossible. The possibility of attack upon problems of teaching and of the adaptation of teaching materials to the needs of pupils comes into view. In short, the teacher as a research worker becomes a possibility. For who could be expected to be interested in conducting an experiment in the teaching of reading if he had not at his disposal a reliable measure of the success of his method? Who would have enthusiasm for establishing the relation between computing ability and ability to solve verbal problems in arithmetic if there were no satisfactory means of measuring either ability?

Accordingly, a consideration of the means now at our disposal for better testing in the schools becomes appropriate in a volume devoted to "Research for Teachers." In what I shall have to say on this subject I shall consider the question of testing from two points of view, namely, (1) general tests devised and standardized for wide usage, such as the intelligence tests and the educational or subject-matter tests; and (2) local tests devised by the teacher under the guidance of improved principles and newer techniques. I shall devote the rest of this chapter to intelligence tests, Chapter IV to educational tests, Chapter V to the combining of these two types of general tests, and Chapter VI to improved local tests or what are commonly called "new-type" tests.

THE FIRST GENERAL INTELLIGENCE TESTS

The Binet-Simon test. Some of us well remember when the first really workable device for measuring intelligence arrived—an importation from France. In the 1905 issue of *L'Année Psychologique,* Binet and Simon, a psychologist and a physician, collaborated in reporting a new instrument for estimating levels of intelligence. Their material consisted of thirty tests. Their purpose as stated at the opening of this epoch-making paper was expressed as follows: "When a child is brought into our presence, our aim is to measure his intellectual capacity in order to learn whether he is normal or retarded." It is, therefore, evident that the Binet-Simon tests were originally devised for the particular purpose of identifying subnormal children—a purpose which accounts for some of the strong points, as well as some of the limitations, of these tests in all their successive editions and revisions.

But what a vista these tests opened up to some of us who fifteen or twenty years ago first realized their possibilities! I recall that on the basis of a home-made translation of the 1908 version I was able to make a considerable impression upon the remote and incredulous authorities at the central office in New York City; and that I thus secured for an elementary school of which I was then principal the establishment of three special classes.

There for me, however, and for so many others, the matter rested. As far as school procedure was concerned, these tests were heroic measures. I remember the time we spent in studying them, and the practice we undertook in our efforts to become expert in handling them. But they always remained something unusual and special, something to be appealed to as a last resort, something to be applied only in a few cases, and certainly something not lightly to be undertaken.

Meanwhile the subject-matter tests were getting under way—the Thorndike Graphometer, the Courtis Arithmetic Tests, Series A, the first form of the Hillegas Composition Scale, and the earliest of Gray's Oral Reading Paragraphs.

These instruments had staunch supporters and determined adversaries. It was clear, however, to the most intelligent supporters and to all the adversaries that something was lacking, although no one seemed able to say what the real deficiency was.

The group intelligence test. And then the group intelligence test, due mainly to the labors of Otis, appeared upon the scene. The Army Test was the first in the field, and its dramatic application to the recruits in the World War is known to everyone. As soon as the Army Test could be released for general use, it was taken up by the schools and colleges of the country and from that time to the present the group intelligence test has been with us.

It is a curious fact that although many persons felt the insufficiency of the subject-matter tests, few seemed at first to realize that what we most needed in order to make our test scores of real worth was intelligence scores to place beside them. Until the advent of the group intelligence test this was practically impossible. The Binet-Simon scale had been used, except to a negligible extent, so exclusively for identifying defectives, it had proved to be of such limited value for normal children above the age of twelve or thirteen, and it had been so difficult to apply that the value of the intelligence measure as a supplement to the measure of ability in school subjects, even if it had been apprehended, could scarcely be realized.

It is now about seven years since the group intelligence tests began to give to school people not occasional measures of intelligence to be treasured and counted "every one apart," but millions of them to be counted together on adding machines and with computing tables, to be tabulated, averaged, transmuted into age scores, and intelligence quotients, and used for administrative purposes.

WHAT THE INTELLIGENCE TESTS MEASURE

Intelligence in the school sense. Not very generally, it seems to me, is the fundamental character of this measurement of intelligence apprehended. If we ask the psychologists, "What is intelligence?" we shall receive a variety of answers. For Ebbinghaus it is the combining activity of the mind. For Wundt it is the ability to give attention or to concentrate. For Binet it is sound judgment—common sense, the ability to make one's way and to get on in the world. For Stern it is the adjustment of thinking to new requirements, the ability to meet new situations.

At first these answers seem so different that a common element appears to be impossible. I think, however, that when we narrow our consideration of the question of intelligence to its manifestation in connection with the school we shall find that all of these answers may be most profitably summarized under the heading, "the ability to learn." The child in his learning must assuredly combine, he must give attention, he must exercise sound judgment—especially in selecting and classifying data—and he must make applications to new requirements and conditions. It, therefore, appears that our best answer to the question, "What is intelligence?" is—so far as school matters are concerned—the ability to learn.

Considered in this light the measurement of intelligence is indeed fundamental for school purposes. Is it not because of intelligence rather than because of physical prowess that man is superior to the brute creation? Neither in strength nor fleetness nor keenness of the senses has he an advantage over the higher forms of animal life. It is his central nervous system which permits him to organize the materials of nature for the purpose of better living. It is by means of his brain and the intelligence which his brain makes possible that he inhibits or redirects his own tendencies to response. It is by this means that he makes plans for the future, that he balances one course of action against another, and chooses the action which is to take place. Just as intelligence distinguishes man from the lower orders of animals, so among men it is intelligence which plays the dominant rôle. Broadly speaking, one individual is better than another as a member of the human family in virtue of mental capacity and achievement.

It is, therefore, reasonable that among civilized races the training of the young for adult living should be addressed primarily to the intelligence. This training consists in giving to the child certain condensed experience which, in our school parlance, we call the curriculum. This selected regimen is intended to work certain changes in him which will fit him for adulthood. In the broad sense of the word he is learning, and since the experiences to which the curriculum subjects him are essentially "new requirements," the strength of his intelligence—that is, of his ability to cope with these new requirements—becomes identified with his ability to learn.

By some such course of reasoning we arrive at the notion already expressed that from the point of view of the school intelligence is ability to learn. It is ability to memorize, to form habits, to develop skill, and to take on attitudes, ideals, and appreciations. It is clear, therefore, that intelligence is the fundamental requisite for school success, and that any plans for teaching the child will be conditioned by the amount and character of this essential human trait. For if there is little intelligence, instruction must be slow of pace, repetitive in treatment, and concrete in content. Whether and to what extent rate of presentation, drill, and abstraction may take place can scarcely be ascertained unless intelligence is measured.

Intellectual inequality. The theory of an earlier day proceeded on the assumption that all children could learn equally well. The doctrine of individual differences was unappreciated. Method books said, in effect, "Do thus and so, and the results will be assured." If the child did not get his lessons he was "bad," and he was beaten. Education, it is now agreed, cannot proceed on the old basis of equality. The doctrine that men are born equal is a grim joke. Neither in physical nor economic status is the doctrine true, nor is it true in respect to mental capacity. It is true, however, in a different and very vital sense. Every child is born equal in the right to become all that his nature permits him to become. With reference to the school this imposes the obligation to study *each* child. First and foremost, it means the study of his intellectual endowment. It then imposes upon the school the obligation to guide and to train him in accordance with his capacity. Thus the modern school becomes highly organized, complex, differentiated. The curriculum is diversified, pupils are classified and reclassified. Some advance rapidly and others slowly. The old school was apparently democratic. It knew but one way of working and it used that way with everybody. The modern school is really democratic; it has many ways of working and it tries to adapt them to the varying needs of pupils. This is a task of far greater magnitude. It involves as a fundamental requirement the study of children, especially in regard to their intelligence or their ability to learn.

THE JUDGMENTS OF TEACHERS

Before the advent of the intelligence tests—particularly those capable of being administered to groups—the only basis on which the intelligence of pupils could be reported was the judgment of those who were acquainted with them. These judgments were and still are highly inaccurate. A parent's estimate of the intelligence of his child is generally regarded as valueless. The teacher's estimate of the actual ability of his pupils, though it may not suffer from the bias of a parent, is nevertheless demonstrably unreliable.

Improved judgments are needed. The chief reason why judgments of intelligence are so erroneous is that there is—or has been until very recently—so little opportunity to check up on the correctness of the estimates. These judgments, however, may and should be improved. It cannot be too often or too emphatically stated that even when all the testing has been done which the wit of man permits, there will still be room for the exercise of good judgment. The greatest present danger in the testing movement is, it seems to me, the danger of making the tests cover all the ground and do all the work. People who are ready enough to weigh evidence and to take thought of conditioning circumstances unaccountably assume the attitude in regard to testing that the results must be unequivocal, accurate to a hair, and utterly reliable as a basis of action. Such a condition is never likely to come to pass, and it is doubtful whether such a condition would be desirable. For my own part, I should look forward with no complacency to the day when in education everything were reduced to the basis of tests and decided by means of formulas.

The expert judgment of teachers will always be in demand. The real problem now is to make that judgment truly expert. If, as I believe to be the case, a part of the shortcomings of educational estimates has been due to the lack of norms against which to check the accuracy of the estimates, then it seems to me that one of the most fruitful uses of the intelligence tests—as well as of other tests—will be their employment to *refine the judgments of teachers.*

Why judgments of intelligence are faulty. As things are at present there are several reasons why teachers' judgments of the mentality of pupils are faulty. In the first place, they are influenced by traits other than intelligence, such as industry, personality, and appearance. The child of pleasing manner is given a higher intelligence rating because of that manner.

In the second place, the age of a pupil influences considerably the judgments of teachers as to his mentality. The old child in a certain class—the ten-year-old child in the second grade, for example—does good work and is rated high in intelligence. As a matter of fact, the situations with which he is called upon to cope are appropriate to children seven or eight years old. His maturity creates a false impression. Very likely he has the mentality of a seven- or eight-year-old child. In that case his intelligence is considerably below normal. Moreover, if he were not retarded intellectually he would probably not be in the second grade.

In the third place, it is claimed by Varner [3] that the age level of the children with whom a teacher is working has a perceptible influence on the reliability of the teacher's estimates of intelligence. He asserts that in general young children are more difficult to rate than old children. For example, when the teachers from whom he had records judged the intelligence of second-grade children, they selected correctly only 22 per cent of the bright children and 53 per cent of the dull children. On the other hand, in the eighth grade, teachers selected correctly 42 per cent of the bright and 63 per cent of the dull children.

In the fourth place, a teacher is profoundly influenced by the general mental level of the group of children with whom he is working. A teacher is often able to identify the bright and dull pupils of a given class without being able to say whether these children are really bright or dull in relation to children in general. The brightest child in a class of low mentality may be no better than normal. The dullest pupil in a class of gifted children may be even better than normal. The expertness of teachers does not often enable them to make the general comparison.

---

[3] Varner, G. F. "Improvement in Rating the Intelligence of Pupils," *Journal of Educational Research,* 8:220-32 (October, 1923).

In the fifth place, there appears to be—at least accordin to Varner—a distinct tendency on the part of teachers t rate the intelligence of pupils too high. This is a curiou commentary upon the generally prevalent grumbling o the part of teachers regarding the stupidity of their pupils It seems, however, that when teachers are called upon t make formal estimates of the mentality of their pupils the tell a different story.

Suggestions for improving judgments of intelligence. Th correction of these different disturbing influences is mainl in the hands of the teachers themselves. Merely to kno that these pitfalls exist is to be aided in avoiding them. good working definition of intelligence as it relates to th school will help in ruling out such non-intellectual trait as industry, affability, docility, and the like. Such a definition, adopted from the Virginia Education Commission has been stated as follows: "By intelligence is meant the degre of brightness of a child. It is not the same as scholarship Think rather of the skill with which a child is able to mee new situations both in school and out. The most intelligent child does not always make the highest marks in school An intelligent child may earn high school marks if he i industrious, if he is physically well, if he is regular in attendance, and if other school conditions are favorable. I these conditions are not favorable he may still be very bright and not earn high marks."

The influence of the age of children may be counteracte if the teacher will bear in mind that the young child in a grade is probably a bright child, while the old child wh may be doing equally good work is almost certainly not s bright. Before teachers estimate the intelligence of children they should have the ages of these children written after their names.

The special difficulty in rating young pupils which Varner believes to exist can scarcely be counteracted by anything less fundamental than experience with and study of little children. Two points, however, may be made in this connection. One is that if it is hard to rate young children, it helps the teacher, and it helps those who may use the teacher's ratings to know this fact—to know in other words that the ratings are less reliable than they would be if they had reference to older children. The second point is that it is precisely with the young children that some of the intelligence tests prove to be most reliable. This is especially true of the Binet-Simon scale.

I have said that teachers find it difficult to relate their particular group of pupils to pupils in general—to tell whether their bright children are really bright from the general point of view and whether their dull children would be dull if all children of their age were taken into account. Undoubtedly this trouble, like the others, may be somewhat diminished by knowing of its existence and by taking thought of pupils in general. It is also suggested that a teacher might well begin by making an estimate of the place of the class among classes of the grade in question—whether it is a good, average, or poor class of its kind. This will help in locating the pupils with reference to pupils in general.

Finally, both the tendency mentioned in the last paragraph and the tendency to rate intelligence too high may be partly corrected by utilizing the method employed in the army during the World War—a method in virtue of which five grades in a given trait were each represented by one or two men whom the rater had actually known. In the same way the teacher may profitably make a child-to-child comparison. Let him think of the brightest child he has known in the grade in question and let him write down the name of this child. Let the brightness of this child stand as A, meaning the highest rating for intelligence. Similarly, let the teacher think of the dullest child of his experience in the particular grade, and let the dullness of this child be represented by E on the scale of intelligence. Then let the intelligence of one or more children from the teacher's experience stand for average intelligence with the rating C associated with it. Finally, let children of the teacher's acquaintance in this grade represent intelligence above average but not of the highest and intelligence below average but not of the lowest. These intelligences may be called B and D respectively.

According to this arrangement we may have a five-point intelligence scale represented by actual concrete instances. Such a scale might look like the accompanying "Scale for Intelligence."

Another way to reduce a bias toward either high or low ratings is to make the ratings conform to a predetermined scheme. Investigators have shown rather conclusively that intelligence is distributed according to what is termed the normal surface of frequency. This means that there is an average group more numerous than any other. These are the children in the C group according to the above scheme.

SCALE FOR INTELLIGENCE.

Rating	Standard Pupils	Pupils of This Class
A	Mary B.	
B	John S. Ethel H.	
C	Henry H. Jane C.	
D	Robert K. Harry L.	
E	Anna R.	

If the normal surface of frequency applies, it likewise means that the number of ratings above and below average are equal and that the farther they are from the average, either up or down, the less frequent they are. This would make the B's and D's of the above classification equally numerous but not so numerous as the C's. It would mean further that there would be the same number of A's as of E's and that this number would be the smallest.

If we apply this normal surface of frequency, we may do so by providing that about 5 per cent of the pupils of a given grade be rated A and that the same per cent be rated E; that 20 per cent be rated B and likewise 20 per cent be rated D; and finally that 50 per cent be rated C. Unless the teacher is convinced that his class is better or poorer in intelligence than the average class of the grade in question he will do well to adhere rather closely to this 5, 20, 50, 20, 5 sequence.

There are in the above paragraphs a few suggestions which may be applied to the rating of pupils for other traits than that of general intelligence. In the first place, a good definition of the trait in question and a caution not to mix it with other ideas will help. In the second place, the factor of age, whether of the individual pupils or of the group, may play a part. Maturity, as has often been pointed out, is in evidence both in emotional and in moral characteristics. It is certainly in evidence in the matter of industry. In the third place, it is probable that the same caution should be made as to the relation between the particular pupils under consideration and pupils in general with respect to the trait with which one is concerned. No doubt, the child-to-child comparison will help here as in the case of general intelligence. Indeed it may help even more. In the fourth place, it is probable that most of the qualities with which we are concerned in school are distributed as biological traits appear to be distributed generally—namely, in a symmetrical manner according to the so-called normal surface of frequency. If this is true the same precaution should be taken to make the percentages conform to the 5, 20, 50, 20, 5 arrangement as in the case of intelligence.

Other judgments than those of intelligence are needed. As we have said before, although the testing movement has gone to considerable length, good judgment on the part of teachers will always be required to supplement objective measures and to take the place of such measures when they are not at hand. I plead for expertness in the making of these judgments whether they be of intelligence, or of emotional or moral traits. Most of the things said above regarding intelligence are applicable to these characteristics. It is precisely because we have fairly accurate measures of intelligence that we have been able to identify at least a few of the causes of error in judging intelligence. Probably our judgments of other personal qualities are vitiated by the same disturbing influences. By attention to these influences, more reliable estimates may doubtless be made. Thus the refining of teachers' judgments applies much more broadly than to intelligence. We realize that success in childhood and in school pursuits as well as in adulthood and in society depends very largely upon emotional characteristics and upon industry, friendliness, and other personal qualities. If the lessons we may learn from the appraisal of intelligence can be carried over so as to refine our judgments of these important human traits, we shall be able to apply educational guidance in a high and worthy sense. We shall make the school something more than a training ground, an institution which will ascertain the capacity and needs of every individual child and regulate its influences upon him according to this expert diagnosis of conditions.

THE YEAR 1919

No consideration of intelligence tests as part of the research instruments available to teachers would be particularly helpful unless it presented some concrete data as to the point of view and conclusions of investigators. The problem of selecting these details from the vast body of literature devoted to the question is by no means easy. I am going to narrow the field by considering the publications of a single year. I have already indicated the remarkable development of intelligence testing which took place immediately after the close of the World War. At that time the Army Test became available and the goodly number of group tests which sprang from it rapidly began to take form. For a year or two the literature on intelligence testing was both voluminous and stimulating. All the larger meanings of the movement were then expressed. Perhaps some of them were overstated and required the more critical formulation of later years. But on the other hand, the more recent literature is comparatively colorless. At its best it deals in details and fine distinctions; at its worst it is merely repetitive. I have, therefore, chosen the year 1919 as *annus mirabilis* in the development of intelligence tests. In that year the enlarged view of their possibilities was first fully expressed.

Two notable books were published on the subject—Terman's *The Intelligence of School Children* and Woodrow's *Brightness and Dullness in Children.* Terman made clear the importance of testing *all* school children, not merely those suspected of being subnormal. He gave the first complete picture of the mentality of school children. He showed the astonishing range of mental ages in a given grade. He pointed out the fact that real retardation is the holding back of capable pupils. He showed that children young for their grades are often the retarded children, because they are bright enough to be in still higher grades. Finally, he contended that bright children were not weaklings nor emotionally unstable.

Woodrow presented many helpful ideas but none more significant than that of the varying rates of development which children exhibit—their chronological, physical, pedagogical, mental, and social ages. The doctrine was enunciated that in these and in other respects the child grows old or matures unequally. Mentally a child may be seven years old and in that case his mental age is seven. Physically the same child may be but five years old and in that case we say his physical age is five. From his contacts with his fellows —the way he "meets people"—we may estimate his social age to be eight. He may have attended school two years and thus have the schooling of an eight-year-old child. In this case his pedagogical age may be said to be eight. Finally, this same child may have actually lived seven and one-half years and this would make his chronological age seven and one-half. This idea of measuring development in terms of ages is an exceedingly fruitful one.

In this same year a legion of magazine articles and bulletin reports on intelligence testing was published. The *Journal of Applied Psychology* was particularly active. Articles appeared in it by the Presseys describing their "cross-out" tests and on "A Comparison of Colored and White Children by Means of a Group Scale of Intelligence"; by Lowell on "A Group Intelligence Scale for Primary Grades," being a collection of exercises from the younger age levels of the Binet-Simon scale arranged for group testing; by Pintner on "Non-Language Group Intelligence Tests"; by Thurstone on "Mental Tests for Prospective Telegraphers"; by Arthur and Woodrow on "An Absolute Intelligence Scale"; by Thorndike on "A Standardized Group Examination of Intelligence Independent of Language." This was by no means all of the articles even in this one magazine on the subject of intelligence tests. Other magazines in this remarkable year, although none of them gave such exclusive attention to intelligence testing as did the recently established *Journal of Applied Psychology*, nevertheless contained in the aggregate far more material on this subject than I have space to consider. A number of these articles reported the application of the army tests to school children and to students in universities. The group intelligence tests which had first become available eighteen or twenty months before were eagerly and widely exploited. Most of the ideas which have since become more or less common property are to be found in the published writings of 1919.

Bright children are not getting a square deal. In the first place, it was shown that capable children were not being identified and properly trained. Terman demonstrated in his book already referred to that the bright children rather than the dullards are the retardates in our schools. Any teacher can verify this in his own class. Get the mental ages of your pupils and you will find a tendency for those who are old mentally—that is, for the capable ones—to be working in grades appropriate for children of lower mental age. On the other hand, you will find that the children whose mental ages are low are graded high with reference to these ages.

Facts useful in guidance. In the second place, educational and vocational guidance received the support of a number of definite and usable facts. For example, the line was pretty clearly drawn between those who could and those who could not be taught reading and arithmetic. If a child's mental age was not more than one-half of his calendar age —in other words, if his intelligence quotient was not more than 50—it was shown that it would be quite useless to try to teach him reading and arithmetic. Children of even higher mentality often fail to profit by the attempts to teach them these subjects. An enormous amount of time and money is no doubt wasted trying to teach children things they can never use. And this, by the way, is an important research topic—one on which teachers may work to advantage.

The application of this idea of the minimum intelligence necessary for success in school work was particularly applied in the high school. It was shown—and it has not since been contradicted—that if pupils' intelligence quotients were below 70 it was useless for them to attempt to study Latin or algebra. It was also shown that children whose intelligence quotients were below 90 were almost certain not to graduate from high school. In connection with this fact it was asserted that one-fourth of the children enrolled in the public schools had less than the degree of mentality required for graduation from high school. With further reference to facts bearing upon educational and vocational guidance, it was shown in the writings of 1919, that even children of a low degree of mentality can be taught to do many useful things. A girl may learn to wash dishes, set table, scrub, and do other forms of housework. She may be taught to do simple sewing, lace making, rug weaving, ironing, and sorting clothes. A boy who will never be older than seven or eight mentally may nevertheless be taught to do housework such as sweeping and polishing, laundry work, teaming, using pick and shovel, brush making, bench work, and net work.

Indeed, feeble-minded folks do some things surprisingly well. Dr. Charters tells the story of the best saleswoman of untrimmed hats in a certain department store. While the simple details of this job annoyed her quicker-witted fellow workers, this woman found these duties sufficiently stimulating for her mentality. By the union of a high degree of interest in a simple task and industry in the pursuit of it she was able to become highly proficient. As we ascend the scale of intelligence, still remaining, however, clearly below the level of normal ability the number of available gainful occupations becomes surprisingly large. The average mentality of street-car motormen and conductors has been reported to be represented by an intelligence quotient of 84. Express company employees have been found to have mental ages as low as ten or eleven years. Policemen and firemen—at least such applicants for those positions as were tested by Terman—do not appear to average up to normal intelligence. The same statement may doubtless be made concerning some other types of workers. It is added, moreover, that in some occupations an intelligence quotient of over 90 is a positive detriment. In the work-a-day world the person whose mental powers exceed those appropriate for his job often becomes an agitator. Moreover, as like as not, he is a poor workman. He takes chances. The job offers no legitimate opportunity for initiative. It lacks interest for the person of superior intelligence so he tries on his own account to put interest into it. He fails to follow the rules of the game and becomes unreliable. If he is a motorman he runs around curves faster than the rules allow, or experiments with various ways of manipulating his controls.

Thus in 1919 the lines were laid down for most of the important researches in the relation of intelligence to vocational aptitude. In the following year Proctor's studies appeared with their illuminating facts as to the mentality required in various occupational groups. The lesson in all this for the teacher is wise guidance of occupational choice on the part of pupils. Proctor showed that 90 per cent of the high-school pupils from whom he had records declared their intention to enter professional callings. Since, however, no more than four per cent of the people in gainful occupations in the United States are in the professional class, it is clear that a vast number of high-school pupils are going to be disappointed. Nevertheless, they are asking for and receiving training appropriate for professional life. One needs to bear in mind not only that the world's work calls for a great many types of service other than professional, but also that the number of gainful occupations, the number of really substantial careers, open to those whose mentality is below normal is surprisingly large. We hear much about blind-alley jobs, but only a little study discloses the fact that society bestows large rewards upon workers in occupations where a strong back and a weak intellect are the prime requisites.

Among the conclusions of research workers which were reached during the first year or two after the group intelligence tests arrived, I have referred, in the first place, to the fact that gifted children are not being identified and suitably trained. In the second place, I have referred to the disclosure of facts vital to educational and vocational guidance.

Correspondence of mental age with school ability. In the third place, it appears to have been definitely established, even in the early years of the expansion of intelligence testing, that a child can do satisfactory work in the grade appropriate to his mental age. On this fact rests in great part the use of the intelligence tests for purposes of grading and classification. The child whose mental age is six can do the work of the first grade. If he falls much below the mental age of six it matters little how many years he may have lived. He is not ready for first-grade work. The notion of promoting children from the kindergarten to the first grade solely on the basis of the time that has elapsed since their date of birth is sharply questioned. Moreover, if children have been properly taught in the first grade they can satisfactorily enter upon second grade work at the mental age of seven years. Similar statements may be made with reference to higher grades.

Bright children are not physically weak. In the fourth place, significant evidence was produced as early as 1919 and confirmed in more recent years that gifted children are quite as likely to be physically strong as are children of normal or inferior intelligence. The notion is rather generally entertained that capable children are likely to be puny. This arises from the old error in virtue of which the age of children is left out of account. Do we not find the bright child in a class with older children? His gifts tend to bring him into social contacts outside of the classroom with those who have lived longer than he. Accordingly, we grow accustomed to seeing the physically immature bright child in the companionship of those who quite evidently outrank him in physical strength and maturity. The thing to do, of course, is to consider him in comparison with children of his own age.

Bright children are socially competent. In the fifth place, the notion that bright children are queer or unsocial was brought into serious question in the publications of 1919 and the evidence adduced since then—by Yates, for example—rather definitely negatives this popular conception. The idea, therefore, that children may not be placed in the grades corresponding to their mental ages because of their moral or emotional incompatability with the older children with whom they will associate is a belief which rests upon no evidence worthy of confidence. These children who have shot ahead mentally will, under ordinary circumstances, be maladjusted one way or the other. We may either leave them with their mental inferiors or put them with their physical superiors. Which maladjustment we prefer depends largely upon the view we take of the seriousness of physical as opposed to mental incompatibility. Of course, the best plan is to place the gifted child with other gifted children who are his equals both mentally and physically. This can only be done in relatively large schools or school systems.

The conservation of intelligence. In the sixth place, in a catalogue of the fruitful ideas that were advanced seven or eight years ago we should not omit that of the conservation of intelligence. In school as well as in the adult world thousands of individuals have greater mentality than their work requires. Such individuals are unhappy, discontented, and idlers of their time. The boy who is compelled to do tasks far under his powers grows up accustomed to no exertion. In school he adapts himself to the plodding pace of the class, develops habits of divided attention, of day-dreaming, and of "getting by" without effort.

There is nothing more fundamental in the way of national resources than the sum total of available intelligence. The assignment of human beings to the jobs of life should be carried out with due regard to the degree of intelligence required for the work in question. To place a person of low intelligence, whether that person is a school child or an adult worker, in a position which requires a higher degree of intelligence is to rob him of the satisfaction of success and to engender the habit of failure. On the other hand, to place a person of high intelligence in a position which calls for lower intelligence tends to weaken the fiber of his moral nature. It is wickedly wasteful of the most precious of human possessions.

Gifted children. Finally, and in close relation to the conservation of intelligence, the knowledge given us in the publications of 1919 was particularly copious on the question of gifted children and their treatment. Terman and Whipple contributed much in this connection. They and others gave evidence in support of the following points: (1) That teachers do not identify gifted children with anything like the certainty with which they select subnormal children; (2) that a far better selection may be made on the basis of intelligence testing than can be made by teachers' judgments; (3) that about one child in ten is "gifted"—namely, has an intelligence quotient of 115 or over; (4) that such children are not only less likely to be physically weak than unselected children but that when given really exacting work to do they are also less likely to be snobbish; (5) that in teaching such children, drill and explanation of new topics may be cut in half; (6) that with them more applications of principles and broader generalizations—that is, larger units of instruction—may be employed; and (7) that accordingly such children—one in ten of them as we find them in school—can do twice as much work as is now required in elementary-school classes.

Such is the record, sketchily reproduced, as I read it for the year 1919. I should add too many pages to this chapter if I should try to show the development and modification of these ideas in more recent years. I am convinced, too, that very little of material value for the teacher would be included. Some of the ideas expressed during the first year or two of group testing have been challenged. Notably

Bagley has combated in the interest of democracy the belief entertained by many of the intelligence testers that the capacity measured by the tests was innate and unalterable. He believes that growth in intelligence is a fact, and he supports his contention with characteristic vigor and ability.

This is an important question; but the dangers which Bagley sees can all be overcome if instead of relying on the intelligence tests to tell the whole story and decide the fate of a child we employ other measures, such as the educational tests, and supplement the test results with other obtainable facts. As I have had occasion to say before, neither the intelligence tests nor the educational tests nor both of these together give us the whole child. Judgment, expert judgment, will still be needed. This judgment, however, will be helped by the tests.

Part of the reason why less important writing is being done on intelligence testing than was the case a few years ago is that these tests have been accepted as a matter of course in many schools. They are being used in a routine way. Our educational periodicals no longer carry articles merely describing some new intelligence test or telling about the giving of an intelligence test to school children. Instead these periodicals are giving space to reports of investigations in which the tests are used as working tools.

Thus the intelligence tests have made possible the study of questions which were not possible before. They have enabled us to predict probable success, to give more accurate advice to pupils, and to classify them better for purposes of instruction. They have permitted us to compare methods of teaching and administrative conditions. We can set one group over against another of the same ability, and we can thus be sure that the result we secure is reasonably free from the influence of the ability of the pupils.

An example will make this clear. In the study which Stevenson has made of the effect of size of class upon teaching, he has first arranged matters (by intelligence testing) so that a large class which is to be compared with a small class has the same degree of pupil ability. If, then, the small class surpasses the large class in achievement, it cannot be said that this was because the pupils in the small class had more ability. The point is that by this device—which would have been practically impossible eight or ten years ago—the disturbing effect of one important influence is ruled out or allowed for.

In your work as a classroom teacher you can make good use of the intelligence-test scores of all your pupils. These scores will tell you, unless there are disturbing influences, what you have a right to expect of your pupils in the way of accomplishment. You would do well, therefore, to secure a few of the best intelligence tests and study them. In doing so you will learn much that I could only give here in an abstract and uninteresting way.

Finally, in addition to the uses you can make of these tests in an instructional way, there is the use you may make of them in studying your pupils and your work with them. This is the research idea. If you take this idea seriously, you will find that you cannot carry it into effect without using the intelligence tests.

It is obvious, painfully so perhaps, that the foregoing chapter does not attempt to give a detailed treatment of intelligence tests. Others have done this far better than I could do it even if sufficient space were at my disposal. I have merely tried to show in a general way the place that intelligence tests ought to assume in your thinking in view of certain results which have been secured and of certain ideas which have been expressed. For anything more far-reaching in the matter of intelligence tests as they apply or may be applied in the schools, I should like to refer you to a very few books which I am sure you will find suggestive. I could make the list much longer, but I have myself been bored many times by the formidable bibliographies which authors have assembled. I have often thought they would be less forbidding—the bibliographies I mean—if they were shorter. This one is short but it is good.

Dickson, V. E. *Mental Tests and the Classroom Teacher.* Yonkers, New York: World Book Company, 1923. 231 pp.

Hines, H. C. *Measuring Intelligence.* Boston: Houghton Mifflin Company, 1923. 146 pp.

Terman, L. M. *The Measurement of Intelligence.* Boston: Houghton Mifflin Company, 1916. 362 pp.

A guide for the use of the Stanford Revision of the Binet-Simon Scale.

Terman, L. M. *The Intelligence of School Children.* Boston: Houghton Mifflin Company, 1919. 317 pp.

Woodrow, Herbert. *Brightness and Dullness in Children.* Philadelphia: J. B. Lippincott Company, 1919. 322 pp.

CHAPTER IV

BETTER TESTING: EDUCATIONAL TESTS

By the term educational tests I mean tests in the school subjects. They are sometimes called pedagogical tests or subject-matter tests. I mean tests of handwriting, reading, arithmetic, algebra, Latin, and all the various subjects of the elementary or secondary curriculum.

The story of the origin and development of educational tests has been too frequently told to justify its repetition here. As I have already said, teachers have probably always tested their work. The difference between the resources at your disposal to-day and those which were available fifteen or twenty years ago is merely the difference between high-grade instruments and low-grade instruments.

THE FIRST HANDWRITING SCALE

A single example will show what is meant by this statement. Teachers used to rate handwriting—and many teachers still do so—by giving it a percentage grade. This pupil's writing was marked 68; that pupil's, 84. What these marks meant nobody knew except that the 84 expressed the teacher's opinion that the second piece of handwriting was better than the first. In 1910 the first handwriting scale made its appearance. It was devised by Thorndike and was sometimes called a graphometer. It consisted of a series of printed samples of handwriting which differed from each other in merit by a uniform amount. The poorest sample was a mere scrawl artificially produced. It was an attempt at writing, but it was wholly without legibility or form or beauty. This sample was called zero. At the other end of the scale was a product of the engraver's art, a specimen as near perfection as could be attained. This was called quality 18. The scale was made up of actual samples of children's writing, and since none was found as poor as quality one, two, or three, the usable part of the scale began with quality four.

Now, if you wish to measure a sample of handwriting you may do so by sliding it along this scale until you reach a specimen on the scale which corresponds in merit to the pupil's sample of writing. You will then give the pupil's sample the number of the corresponding specimen on the scale. This, then, is the quality or merit of the pupil's sample.

This sort of rating is very much more satisfactory than a rating which is based entirely on personal judgment. I am reminded of the method used at creameries for measuring the cleanness of milk. The same quantity of the milk is drawn from each can and filtered through standard filter paper. The discoloration of the filter paper is then matched against a series of colors which have been previously standardized. Thus a can of milk may be said to rate at one or five or eight in point of cleanness. I am told that before the introduction of this system there was always difficulty in enforcing the rules of cleanness in the handling of milk. An argument could always be made as to whether or not the milk was sufficiently clean. When, however, the plan I have just described was introduced the milk told its own story, and the rating was at least visible and tangible.

Some have objected that quality in handwriting is too intangible to measure—that an impressionistic or judgment rating is of little value. Some have even denied the possibility of it. The only point we would make in this connection is that the old method of assigning percentage grades was quite as impressionistic, quite as much an estimate of general merit as that which the scale provides. Very few practical teachers have ever raised, in all the controversy that has taken place concerning the use of tests and scales, the question of their own ability to judge the quality of such products as samples of handwriting or drawing or English composition.

It is true that in using the handwriting scale and other scales of similar nature the teacher's personal judgment plays an important part. There is, however, this difference. The judgment is always capable of being checked against something objective, something visible, something tangible, something you can carry around in your pocket and not merely in your head, something you can show to people—to other teachers, to supervisors, to parents and to pupils themselves—and finally, something which has the same meaning to one person that it has to another. If I say that this pupil's handwriting is of quality 12 on the Thorndike scale, it is entirely possible for another to verify the accuracy of my judgment. But if I say that the same piece of handwriting is worth 65 per cent, nobody knows what I mean.

What I have said as to the superiority of testing by the use of the handwriting scale may be said in general of the superiority of testing by means of other measuring instruments. No modern teacher will be without them as a means for securing "better testing." Every such teacher owes it to himself to obtain price lists and samples of tests from the publishers of these materials. Almost all the tests now on the market may be secured from one of three publishers, namely, World Book Company, Yonkers, New York; The Public School Publishing Company, Bloomington, Illinois; and the Bureau of Publications of Teachers College, Columbia University, New York City.

SOME TENDENCIES IN EDUCATIONAL TESTING

General to specific. I now propose to show what I think are the most significant tendencies exhibited by the test movement. In the first place, there has been a definite trend from the general to the specific. The first handwriting scales had reference merely to general merit. Somewhat later general merit began to be analyzed, and we had analytical instruments. The Ayres scale distinguished between vertical, semi-slant, and full slant writing and was, therefore, more analytical than the Thorndike scale for general merit which preceded it. Then came the work of Freeman and Gray, in which the quality of handwriting was still further analyzed into such matters as uniformity of slant, alignment, line quality, and spacing.

The same development from the general to the specific took place in connection with the measuring of English composition. The Hillegas scale gave a series of samples of composition to which were attached numerical values. The idea, in other words, was much the same as in the Thorndike handwriting scale. Straightway, however, merit in English composition began to be analyzed, and scales began to be produced in conformity with the analysis. Several samples were exhibited at each step on the scale. The Harvard-Newton device offered separate scales for narration, description, and exposition; and the Willing scale distinguished "story value" from merit in matters of form.

In arithmetic—to mention one more subject—column addition as measured by the Courtis Tests (Series B) is a general ability. It involves the handling of nine three-place numbers in each example. But addition, as Courtis was himself one of the first to point out, involves a great many specific skills. Accordingly, tests have appeared which address themselves to at least some of these skills. The analytical tests—those which concern themselves with the more specific abilities—came to be called diagnostic tests, because they could be used in finding out why the general result was unsatisfactory. Thus Monroe's Diagnostic Test in Arithmetic would go further than merely to show a pupil's general adding ability. It would show whether or not he knew the fundamental addition combinations and if he did not know them which ones were missing from his repertoire. It would show whether he had been able satisfactorily to make the transfer from fundamental combinations to higher decade[1] combinations as these are used in column addition. It would show by columns of various lengths how far the child was able to sustain his attention. It would show by simple columns of no great length whether he was able to carry.

---

[1] Here I adopt the terminology of Osburn. A higher decade combination is one in which a one-place number is added to a two-place number; for example, 12 + 6, 37 + 5, etc.

Thus a diagnostic test affords a criticism of the general test. The latter is often called a *survey* test. The point we are making in this connection is that all the original tests were survey tests and that among the later tests there are many analytical ones. Survey tests were useful and they still have important uses. By them we determine whether our classes are up to grade and how they compare with other classes. Some of them are also useful in reference to individual pupils. We find out by a general scale or test what a pupil's ability is. Then we measure the various subordinate abilities to find out the causes of such lack of general ability as the first testing may have revealed.

Simplification. In the second place, there has been a decided trend among test makers in the direction of simplifying tests. The influence of some of the techniques employed in the group intelligence tests began to be felt as soon as these tests made their appearance. For example, instead of asking a pupil to work a problem in arithmetic and set down the answer, an author of a test may give him four or five choices of answers, one of which is correct. The pupil is merely required to underline the correct answer. Again, in a geography test three or four reasons may be given for a phenomenon with the requirement that the pupil select the best reason. He may exhibit his knowledge of English construction not by writing an essay but by filling in blanks with the proper words. He may show his ability to read a given sentence by crossing out the word which does not belong to it. His knowledge of facts in almost any subject may be exhibited by the accuracy with which he detects true and false statements. In all these cases the answer is indicated by a word or a letter or a mark and these words, letters, or marks are made in a definite place. This exact location permits a mechanical device to be used in rating the papers. Either a stencil may be made with apertures exactly over the location of the correct response or, by arranging the test items so that the responses come in columns, a matching device may be used. Thus the rating of pupils' answer papers is made much more accurate and is accomplished much more rapidly.

One reason why simplicity is necessary. This development toward simplicity in giving and scoring tests is necessary. In my opinion it is *the teacher* and not the supervisor who should give the tests. These instruments are applied because it is believed that teaching will be improved through better understanding of the abilities and disabilities of pupils. It is clear, however, that the tests themselves teach very little—that is, they teach the children very little. It is a mistake to think of them as teaching devices. They are measuring instruments, and they fulfill their purpose when they register accurately the conditions of the moment. If, therefore, teaching is to be improved, it is the teacher who must accomplish this; and if the teaching is to be improved through the light thrown upon conditions by the tests, the teacher, and not somebody else, should give the tests. The maximum information derivable for teaching purposes and convertible into teaching efficiency will be secured when the agent of instruction, namely, the teacher, handles the tests and rates them.

While I am on this topic, I want to urge the giving of tests by teachers on another ground—the ground, namely, that it is the teacher's right to do so. It is your job and you should insist on exercising it. I do not mean that tests given for administrative purposes are not properly given by some one in supervisory authority. If the purpose is to check upon your work—and every good teacher should welcome this—then some one else must do the testing. If the purpose is the general survey purpose—comparison of class with class, school with school, etc.—either the teacher or a representative of the superintendent's office may give the tests. In this case, if the teacher does the testing, he is in fact *acting for* the superintendent. But if testing is to form the basis of better teaching—and it generally should—then the teacher must be the tester. If you permit any one else to do this work, you weaken your authority and your position. Moreover, the teaching which is supposed to be benefited will be feebly affected.

This means that you must *be able* to give the tests and to give them well. Remember that it is far more important that a job be done, and done right, than that a particular person should do it. If you are unequal to this task or indifferent to it, some one else should, and very likely will, take your place.

The testing program. In the third place, just as we have a movement from the general to the specific in testing, so we have a movement from the general to the still more general. Instead of giving an arithmetic test or a reading test or a spelling test, we are giving batteries of tests. In many school systems a test in a single subject has been given over a series of years. "Campaign" methods have been employed to stimulate both teachers and pupils in the particular subject. Such a school is bound to be developed in a one-sided way. It is better to adopt a well-rounded testing program. By such a program the general level of school abilities will be improved, and one set of abilities will not be emphasized at the expense of others.

A simple example of the battery of tests is the so-called "Second-Grade Achievement Tests," by Dr. Luella C. Pressey. This test offers in a single leaflet a close approximation to the course of study of the second grade. It measures achievement in spelling, arithmetic, reading, and vocabulary as applicable to that grade. Another example of the battery idea is the Lippincott-Chapman Classroom Products Survey Test. This has the advantage of being applicable to a series of grades and, therefore, of affording data which can be used for purposes of regrading pupils. This battery consists of tests in (1) arithmetic fundamentals, (2) reading a continuous passage, (3) arithmetic problems, and (4) reading selections.

The Illinois Examination is a battery with a little different idea. It comprises tests of the two most fundamental elementary-school subjects, namely, reading and arithmetic, together with an intelligence test. The significance of combining the intelligence test with the subject-matter tests will be discussed later. One other battery should be mentioned, namely, the Stanford Achievement Test. It represents the test maker at about the highest point he has reached. As the name implies, it has to do entirely with the achievement of pupils in the different school subjects. Its usefulness will, in my judgment, be considerably enhanced if a good intelligence test is given with it. For grades above the third, this battery of tests includes not only reading and arithmetic (two or three phases of each) but also nature study, history and literature, language usage and dictation.

Manufacture and sale of tests. In the fourth place, a development has taken place in the production and distribution of test materials. Up to 1919 tests were published for the most part by their authors. They were printed by anyone whom the author could get to undertake the job. The author usually sold the tests either personally or through an institution with which he might be connected. He was generally a research man and after the test had been brought out he was likely to become interested in something else. In other words, the marketing of his tests was unlikely to occupy much of his attention. The result was that the testing instruments were printed in all kinds of ways—many of them poorly—and on various sizes and shapes of paper. When the first edition was exhausted, it was often difficult to bring out a second edition. Those who ordered tests frequently had to wait long for deliveries. Moreover, if they wanted several tests they were obliged to order them from several different persons or institutions.

All this is very much better handled now. In 1919 the commercial houses began to take over the manufacture and distribution of tests and scales. The demand for these instruments had grown so great that the production facilities at the disposal of individual authors could not keep up with it. The result has been an improvement in every particular except price. The publishers have secured expert service, and this service is utilized not only in shaping up material according to the best prevailing practice but also in giving advice to those who use the tests. The numerous sizes of tests have been reduced practically to two. Directions and lists of acceptable answers have been printed and distributed with the tests. Stencils and tables for handling the scores, record sheets, and the norms pertaining to the tests have likewise been assembled and issued in acceptable form. Moreover, one no longer has to send to a half dozen places in order to obtain as many different tests. On the whole, therefore, the taking over of this business by the commercial houses has benefited the testing movement.

Critical matters. In the fifth place, with the development of the test movement the demand for an ever-increasing critical evaluation of the tests as measuring instruments has become apparent. The test-makers in seeking to meet this demand have distinguished several characteristics which a good test should exemplify. Chief among these is the *validity* of the test. This term "validity" is a technical word representing the idea of the general worth or merit of the test as a measuring instrument applicable to the thing with which it is concerned. According to this conception, a reading test has high validity when it measures the thing we call ability to read. Clearly, in order to establish the validity of a test it is necessary to have some measure of the trait in question with which the test's results may be compared. This is a good deal like lifting one's self by one's boot straps. We devise a reading test and then we must have a test of reading with which to compare it. Progress in validation, however, may be made by securing for a limited number of pupils an exhaustive record of their reading ability —a record to which not only several standardized tests of reading may contribute but also informal tests, results of interviews, and opinions of teachers. By dint of much labor a very acceptable estimate of the reading ability of each of these children may thus be secured. These data, with the basis upon which they rest, may be called *the criterion.* This criterion is a standard of reference for the new test whose validity we are about to determine.

Having thus set up a criterion, we now determine the validity of the test or scale with which we are concerned by finding out the extent to which it tells the same story as the criterion. The criterion, as you will readily see, has been secured at great effort. It may have required many hundreds of hours of labor to attain it. If now we give the new test to the same children and if we secure the same result, or nearly the same result, as we did by the laborious method of setting up the criterion, then our test is valid. It is a good reading test. It measures the thing it purports to measure. Moreover, it can be given quickly and cheaply in comparison with the methods thought to be necessary in setting up the criterion.

A test may be valid within limits and have very little validity outside of these limits. The author, in simple justice, should state accurately the range of educational or mental functions for which the test is claimed to be valid. A certain instrument may be called a history test, but if it relates to the Colonial period it can be valid only with reference to that period. Similarly a set of exercises may be called a reading test, yet it may address itself only to one or two phases of reading ability, such as oral reading or the ability to read a paragraph silently and pick out its main idea, or the ability to give the meaning of words. Again, a test may be valid only for children of certain ages or grades. You are entitled to know what a test is good for, and the most important consideration is its validity.

Another way in which authors have sought to improve their tests is in respect to *reliability*. Like validity, reliability has become a technical term. It means the accuracy with which the test measures. If we give a test twice to the same group of pupils—or, better still, if we give two equally difficult forms of the same test—and obtain the same or nearly the same scores on these two occasions, we may say that the test is reliable. It yields the same result each time it is administered to the same group. Since no test can very well measure what it attempts to measure unless it does so with accuracy, we may say that any test which is highly valid is also highly reliable. The opposite, however, is not necessarily true. A test upon repeated use with the same pupils may yield the same results without these results being truly indicative of the ability which we are seeking to measure. This has sometimes turned out to be the case. For example, the Trabue Language Tests yield much the same score with the same individuals on different occasions. They do not, however, afford a good measure of language ability, unless we give a very special definition to the term "language ability."

Both validity and reliability are essential characteristics of a good test, but validity takes precedence over reliability. A test must measure the thing it purports to measure, or its accuracy is worthless. No one is interested in the precision with which a test measures something other than what it is intended to measure.

Transmuting test scores. In the sixth place, the users of tests, as well as the makers of them, have derived from the test scores certain measures which have greater value for school purposes than have the test scores themselves. One of the reasons why the test scores as at first obtained —the raw scores, as they are called—are of limited value is that the units in which they are expressed have such different meanings and such different values. For example, *one* in the Courtis Arithmetic Test in addition means an example of 27 digits summed correctly; in the Monroe Silent Reading Test *one* means a word read in a minute; in the Ayres Spelling Scale, *one* means a word spelled correctly; while in the Thorndike Handwriting Scale *one* means one-eighteenth of the range of ability in handwriting. The result is that a raw score has no meaning except with reference to the particular test and the units associated with it. A score of 25 would be very high as a comprehension score on the Monroe Silent Reading Test. It would be a very low score in rate of reading according to the same test. It would be equal to about the fifth-grade standard if it was obtained from the Buckingham-Stevenson Place Geography Test (United States). It would mean no more than third-grade ability if it referred to the Illinois General Intelligence Scale and no more than second-grade ability if it referred to the Ayres Handwriting Scale.

Grade scores. It is clear that if a given number may have all these different meanings, we need some way to bring out these meanings. I have suggested the *grade score* as a basis for rendering raw scores comparable. *One* will then be the step from one school grade to the next and all test scores, no matter what their original units may be, will be converted into grade units. This method, by affording a common unit, at once provides a universal basis of scoring. We are no longer obliged to guess at the meaning of a score of 21 on Test II of Monroe's Silent Reading Test or the meaning of a score of 60 on the Ayres Handwriting Scale. We say that the one is 7 and the other 8, meaning that the one corresponds to the beginning of Grade VII and the other to the beginning of Grade VIII. Thus we may transmute raw scores of different and often obscure meanings into scores having a common, an obvious, and a practical meaning.

All we need for this purpose is the grade norms published with each test together with a statement of the time of year at which the testing which led to the norms was done. The time of year is required in order that we may adjust the norms to what they would be at the beginning of each grade. Consider Table XI, on the following page, which has been constructed to show for a few tests the norms adjusted to the beginning of each grade.

Now consider the case of a pupil who scores 9 in the addition part of the Courtis Test, 60 on the Buckingham Problem Test, 175 in rate and 92 in comprehension on the

Courtis Silent Reading Test, 50 in rate and 60 in quality on the Ayres Handwriting Scale, and 75 in story value and 5 in form on the Willing Composition Scale. By using Table XI we may convert these eight measures into grade scores. Let us consider the school year as made up of ten months. Accordingly, the score of 9 in addition, lying as it does 5/6 of the way from 8, the score for the beginning of Grade V, toward 9.2, the score for the beginning of Grade VI, may be recorded as 5.8. This means that the score of 9 places the child in the eighth month of the fifth school year. From another point of view, this score, 5.8, means fifth grade and 8/10 of the way toward sixth grade. Similarly the raw score of 60 on the Buckingham Problem Test yields a grade score of 5.7, and the raw score of 175 in rate on the Courtis Silent Reading Test yields a grade score of 6.3. The complete transversion of the eight scores to which we are referring is given in Table XII. I suggest that you verify the figures in the last column of this table.

TABLE XI. GRADE NORMS ON CERTAIN TESTS ADJUSTED TO THE BEGINNING OF EACH YEAR

Grade	Courtis Arithmetic, Attempts in Addition	Buckingham Problem Test	Courtis Silent Reading		Ayres Handwriting		Willing Composition	
			Rate	Comprehension	Rate	Quality	Story Value	Form (Errors)
VIII	11.2	82	...	..	77	60	77.8	4.4
VII	10.3	68	191	95	73	56	75.3	5.1
VI	9.2	63	168	93	67	52	71.1	6.3
V	8.0	53	145	89	59	48	63.4	8.8
IV	6.2	42	113	78	48	44	50.1	14.6
III	...	32	84	59	36	40	...	...

Observe now that since all the grade scores mean the same thing, we can obtain their average. We could not do this when the units were all different. In getting an average of several items, one has to add the items and, as everybody knows, one cannot add things unless they have the same units. For example, one cannot add 9 addition examples to 60 units of problem-solving ability, nor these in turn to 175 words read per minute. But grade scores in as many tests as are provided with standards may be combined into an average or general grade score. The pupil whose scores are given in Table XII earned an average grade score of 6.2. This means that the pupil performed, on the average, somewhat better than might be expected of a pupil beginning the sixth grade.[2] If we desire to look more closely into the

TABLE XII. RAW SCORES AND GRADE SCORE EQUIVALENTS

Test	Raw Score	Grade Score
Courtis Addition	9	5.8
Buckingham Problems.	60	5.7
Courtis Silent Reading:		
Rate	175	6.3
Comprehension	92	5.8
Ayres Handwriting:		
Rate	50	4.2
Quality	60	8.0
Willing Composition:		
Story value	75	6.9
Form	5	7.1
Average (unweighted).	xxxx	6.2

---

[2] The assumption here is that the abilities covered by the different tests are equally important. This is probably not in reality the case. Very likely rate of handwriting is of less importance than story value in composition. Both may be less important than comprehension of silent reading. The reader may weight the grade scores according to his judgment. When this is done an average may be obtained. This idea of weighting is quite independent of the idea of transmutation.

performance of the pupil we may compare his success in one subject with his success in another by inspecting the grade scores. We cannot do it by inspecting the raw scores. Notice that a raw score of 5 in form on the Willing Composition Scale is a better score than 92 in comprehension on the Courtis Silent Reading Test, and better than 175 in rate on the same test. The grade scores bring out the true condition. According to them this pupil is particularly slow in handwriting, but the quality of his writing is exceedingly high. His composition work is especially good and his arithmetic work fair. In reading his rate is more meritorious than his comprehension.

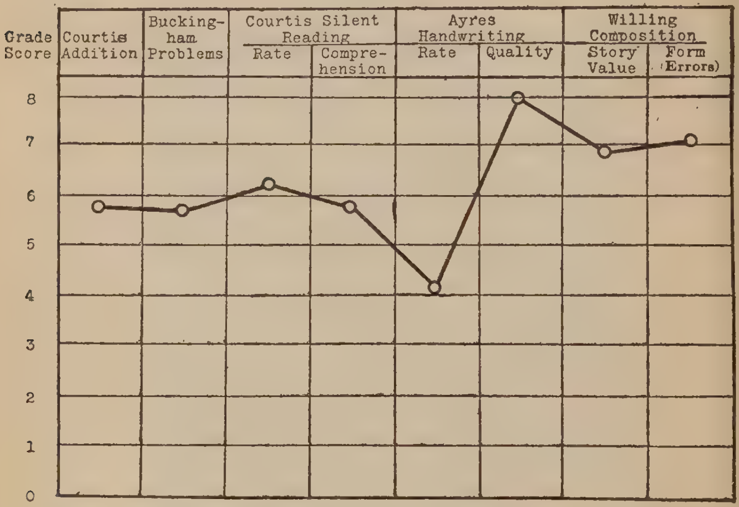

FIGURE 3. PUPIL'S GRAPH, DATA FROM TABLE XII

Many attempts have been made by various writers to show test records in graphic form. The grade score tranversion lends itself very well to this sort of thing. Consider, for example, Figure 3 which exhibits a graph of the same pupil's scores.

Age scores. I have illustrated by the grade score one type of change which we may make in test scores in order to give them greater usefulness. There are many other ways of accomplishing this purpose. One of these ways is to change the test scores into *age scores.* This is done by somewhat the same process as I have described above with reference to grade scores. For example, a rate score of 160 on the Monroe Silent Reading Test (Revised) is the score appropriate to a pupil just twelve years of age. A score of 44 on the General Survey Test in Arithmetic by the same author is likewise typical of a twelve-year-old child. Accordingly, the changing of these test scores to the age-score basis will represent them both by twelve.

Theoretically the age score is superior to the grade score. It is also more useful when we wish to combine educational tests with intelligence tests. This arises from the fact that the makers of intelligence tests have, for the most part, converted their scores into ages, that is, into "mental ages." There are, however, certain practical difficulties in connection with the age score. In the first place, an accurate age score is hard to obtain. This is especially true for school children below the age of nine and above the age of thirteen. Not all of the former are in schools, and many of those who are in school have not yet reached the grades to which educational tests are usually applied. They are likely to be stupid children and if they were tested they would bring down the scores. Since they are not tested the scores obtained for these ages are too high. They are obtained from the brighter children who are in school and who have reached the grade to which the test is given.

On the other hand, age scores for children above the age of thirteen are too low. Indeed some of the brightest twelve- and thirteen-year-old children are through the elementary school, and a goodly proportion of the more competent fourteen-year-old children have certainly finished school. It is very difficult to give the tests to children outside the elementary school. It is not sufficient to follow the pupils into the high school for a great many of them, after leaving the elementary school, do not enter high school. Accordingly, since, in the testing, many of the more capable children must of necessity be left out of account, the age standards derived from the rest of the pupils are manifestly too low.

No small effort has been made by investigators to overcome this difficulty of securing accurate age standards in educational tests. The result, however, has not been wholly satisfactory. It has been necessary for them to make certain assumptions as to growth in ability in the earlier and later years of the elementary-school period, and these assumptions hold true only in a rough way.

There is another practical difficulty. The uses to which you as teachers will want to put these tests are not directly served by the age standards. What you want to know concerning a pupil is whether or not he is up to grade. If he is not, you want to know to what grade his ability corresponds. One of the major uses of tests is for the better grading of pupils. If raw scores are transmuted into age scores, then age scores must in turn be transmuted into grade scores before the desired objective is reached. Each added step makes the final determination less accurate. For example, if you find that one of your pupils reads at the rate of 160 words a minute and if you are given the fact that this is equivalent to the average performance of a twelve-year-old child you must still answer the question, To what grade does the average twelve-year-old child belong? Now a twelve-year-old child may appropriately be either in the sixth or seventh grade. If he is exactly twelve he is, so to speak, in-between—something like the upper sixth grade or the lower seventh. But this is only an average. Many twelve-year-old children are in the fourth or fifth grade, and some have passed through the eighth. You have already dealt with one average when you have changed the test scores into age scores. This is, however, unavoidable; and, I may add, one has to do it in the changing of test scores to grade scores. But why should another average be piled on the first? It would be better—if you want to interpret test scores in terms of grades—to go at once to the grade equivalents. You will then be using but one average, namely, the average score appropriate to the beginning of a grade.

The Detroit plan. A very sensible way of transmuting raw scores into measures and suggestions which the teacher can use was devised by Dr. S. A. Courtis for use in Detroit. According to this plan the teacher divides the pupils into five groups according to their test scores. These five groups are adjusted to the grade in question. Accordingly, the Detroit plan is essentially a grade-score transmutation. It goes further, however, for it has something to say about the treatment of pupils above and below standard. Suppose, for example, a standardized spelling test of 50 words is given to a certain sixth grade consisting of 34 pupils. According to the Detroit plan, the pupils who had perfect papers would be put in Group I; those who had one to three words wrong, in Group II; those who had four to nine words wrong, in Group III; those who had ten to eighteen words wrong, in Group IV; and those who had more than eighteen words wrong, in Group V. Now suppose that in this class of 34 pupils the number of pupils in each group was as follows:

Group	Number of Pupils
I	9
II	12
III	5
IV	4
V	4

A system of weights is now adopted, Group I being given a weight of 10; Group II, 9; Group III, 7; Group IV, 4; and Group V, 0. If it were not desirable to compare one class with another, each pupil in Group I might thus earn 10 points for the class. Each pupil in Group II might similarly earn 9 points, etc. But the class-to-class comparison is important. Accordingly, the number of pupils in each group is converted into a per cent of the entire class. Thus in the sixth grade referred to, the nine pupils in Group I constitute 26 per cent of the entire class of thirty-four. The twelve pupils in Group II are 35 per cent of the class; and so on.

		No.	%	Weight	Points
Group I	CHILDREN OF STANDARD ABILITY (0 mistakes, or perfect papers) Thomas Jacob Alice H. Paul Ella George Helen Mary Charles P.	9	26	10	260
Group II	CHILDREN FOR WHOM REGULAR WORK WILL FURNISH SUFFICIENT DRILL 1 2 3 Harold Horace Adam Kate B. James Inez Ruth Gertrude John J. William R. Edith William A.	12	35	9	315
Group III	CHILDREN IN NEED OF THOROUGH DRILL 4 6 8 Charles K. Patrick Alice M. 5 7 9 Elizabeth Jane	5	15	7	105
Group IV	CHILDREN WHO NEED SPECIAL ATTENTION AND EXTRA DRILL 10 13 16 Kate L. John S. 11 14 17 Mildred 12 15 18 Florence	4	12	4	48
Group V	CHILDREN FOR WHOM SOME SPECIAL ADJUSTMENT OF WORK MUST BE MADE (Record here all scores larger than those above) Henry 19 wrong Roger 24 wrong Mabel 19 wrong Orville 26 wrong	4	12	0	0
	Totals	34	100	xx	728

FIGURE 4. THE DETROIT CLASS RECORD SHEET IN SPELLING

Finally, the per cent of the class in Group I is multiplied by the weight attached to Group I, namely, 10. This yields 260 points toward the class score. Similarly the 35 per cent of the pupils in Group II is multiplied by 9, the weight attached to Group II; and this yields 315 toward the class score. The situation is shown in full in Figure 4, page 189.

Observe that according to Figure 4, Thomas, Paul, Helen, and six other pupils are in Group I. They had perfect papers. Note also that in Group II, Harold and three others had one word wrong, that Horace and four others had two words wrong, and that Adam and two others had three words wrong. The notation in connection with this group is: "children for whom regular work will furnish sufficient drill." In other words, the teacher is not expected to do individual work with these children. The pupils in Group III got from four to nine words wrong. Charles K. had four wrong; Patrick had six wrong; and the other three had seven or eight words wrong. These children are "in need of thorough drill." Mildred, Kate L., Florence, and John S., in Group IV, had eleven, thirteen, fifteen, and sixteen words wrong respectively. According to the record sheet they "need special attention and extra drill." Finally, the four children in Group V require some special adjustment of work.

According to Figure 4, the class score on this spelling test was 728. The highest possible score would be 1,000. This would occur when all the children—that is, 100 per cent of them—were in Group I.

Not only does the Detroit plan present interesting suggestions for grouping children and for applying remedies, but it also permits a comparison among classes and between

	Children of Standard Ability			No.	%	Weight	Points
Group I	R 65—84 Q 80 & up	R 85—94 Q 70 & up	R 95 or better Q 65 & up			10	
	Children for Whom Regular Work Will Furnish Sufficient Drill						
Group II	R 55—74 Q 70 & up	R 75—84 Q 60 & up	R 85 or better Q 55 & up			9	
	Children in Need of Thorough Drill						
Group III	R 45—64 Q 55 & up	R 65—74 Q 45 & up	R 75 or better Q 40 & up			7	
	Children Who Need Special Attention and Extra Drill						
Group IV	R 30—54 Q 45 & up	R 55—64 Q 30 & up	R 65 or better Q 25 & up			4	
	Children for Whom Some Special Adjustment of Work Must Be Made						
Group V	R less than 30 Q any	R 30—54 Q 25—44	R any Q less than 25			0	
	Total						

FIGURE 5. DETROIT CLASS RECORD SHEET IN HANDWRITING

subjects. Figure 5 shows the way the record sheet is adjusted to the reporting of handwriting. In this record shee the reference is to the Ayres Handwriting Scale. R stands for rate—that is, letters written per minute—and Q for quality or merit as measured by the Ayres scale. Observe that a pupil may have his name entered in Group I under three conditions. If his rate of writing is between 65 and 84 letters per minute with a quality of 80, he appears in the first division of Group I. If his rate is more rapid his quality need not be so high, and he appears in the second or third division of Group I. The names of the children are written under the particular rate and quality to which they belong.

Notice that the teacher's attention is called to the needs of the children who fall into each group. Finally, note that the weights attached to each group are *the same as in the case of spelling*. If, therefore, the same per cents of the children fall in each group the class scores will be the same. Again the greatest possible score is 1,000. This means not only that classes may be compared in the same subject but also that comparison may be made between subjects. A class score of 728 has the same meaning for any subject or any grade.

Limitation of training. Another idea which has been developed along with the measurement movement is the idea of the limitation of training. Consider the subject of handwriting. Here is an important skill—important in the sense that society demands it universally. The school cannot evade this demand. Yet it is certain that a pupil may spend too much time in practicing writing and be expected to show a proficiency in it which is out of all relation to its social value. This is precisely what happened in earlier days. Writing was taught with great effort and with great expenditure of time. Moreover, every child, no matter how well he could write, continued to try to improve his writing to an indefinite degree.

When, however, the writing scales came upon the scene, the question was raised, How well does a child *need* to write? It was found that writing of Quality 12 on the Thorndike scale (about Quality 65 on the Ayres scale) was as legible as handwriting can be. Now, Quality 12 is not as pretty as Quality 13 and much less so than Quality 16 or 17. But the amount of school time necessary to secure, even from a few pupils, writing of these higher qualities was very great without producing a result which was any more easily read. If the purpose of handwriting is to be read and if, as may reasonably be maintained, other matters of quality are relatively unimportant, then it becomes the duty of the school to see that the desired quality is secured. But when this quality *is* secured, additional time spent in securing higher quality is not economically spent. It is better to devote time and effort to something else more valuable for the child than added quality of handwriting.

The handwriting scales have also been used to find out from people who employ writers the degree of merit in penmanship which they require. It has been found that for almost every purpose a quality of 60 on the Ayres scale, about 11 on the Thorndike scale, is all that society is demanding. In other words, society does not appear to be requiring writing even of maximum legibility. If, then, through the instrumentality of the handwriting scales we find that civil-service examiners, mail-order houses, and people in general in handling their personal affairs require no more than a specific degree of merit, why should the school be called upon laboriously to develop a greater degree of merit? Why the endless copy-book practice or, more recently, the equally endless push-and-pull exercises and rhythmic formation of compact ovals?

Here then we have a project; or, as the Dalton school calls it, a contract. The pupil with full knowledge of what various qualities of handwriting are—for he has before him the handwriting scale—contracts to attain a perfectly definite goal, which in respect to quality of handwriting meets every reasonable need. Quality 60 on the Ayres scale is reached by the average pupil toward the end of the sixth grade or at the beginning of the seventh. Unquestionably, with this objective definitely in view, the required quality can be secured even earlier in the school course.

As in the case of handwriting, so in the case of arithmetic, with reliable instruments of measurement we may with some certainty ascertain the degree of skill required to meet the needs of life. I am sure that many schools are continuing to spend time on arithmetic beyond any need which the pupils are likely to have for the subject. Professor Wilson's study comes to mind. He showed that adults have relatively little need for performing operations other than those with whole numbers and fractions of small denominators. Even within the realm of whole numbers, a high degree of skill such as is manifested in rapid calculation is not ordinarily demanded. Now that we have standards in arithmetic, it is entirely possible to attain an accurate statement of the degree of skill customarily demanded by society. The present eighth-grade standard according to the Courtis Arithmetic Test includes the adding of twelve examples of nine three-digit figures in eight minutes with 85 per cent accuracy. This ability is secured by the average pupil. There is nothing to indicate that the school is justified in attempting to develop greater skill than this in addition. When pupils have met this standard they may rightly be expected to devote their time to matters more important than a higher degree of adding ability.

Thus the standards based on educational tests reduce effort and release energy. They permit a pupil to devote more time to work of greater value to him than additional practice in writing or drill in arithmetic. What I have just said offers an illuminating commentary upon the idea that standards formalize education. It turns out that at least one use to which standards can be put is to *limit* training in the formal subjects and to release time and energy for higher uses.

The limitation of training, however, does not always apply. There are some lines of school work concerning which our standards for the various grades or even for social uses do not permit us to assign an upper limit of desirability. The standard in reading, for example, for elementary-school graduates on material of about sixth-grade difficulty is approximately 250 words per minute with an 80 per cent accuracy of reproduction. Nevertheless, we are justified in attempting to raise these standards. The standards are, in this case, points of departure rather than stopping places.

Moreover, there is abundant evidence to show that raising these standards is not difficult. Rapid reading with a high degree of comprehension is so distinctly an advantage both to the child and to the adult that we cannot, as in the case of handwriting, set a limit beyond which it is undesirable to go. If we could teach children to read twice as fast as they now do and at the same time teach them to comprehend their reading better than they now do, we should increase enormously their success in high school and their resourcefulness in life.

CONCLUSION

I have not sought in this chapter to give many details about educational tests. In Chapters VII, VIII, and IX I shall have something much more specific to say about the use you can make of tests. There are a number of books which present much more information about particular tests than I can give in the space at my disposal. You should have a few books on testing. You should also obtain samples of the tests which cover your field of interest. Write to the World Book Company, Yonkers, New York, and to the Public School Publishing Company, Bloomington, Illinois, for catalogues. From these make your selection. If you are in doubt write a letter of inquiry to either of these companies. Both of them maintain expert service and they will gladly answer your questions. For the Courtis tests address Courtis Standard Tests, 1807 East Grand Boulevard, Detroit, Michigan. For the tests published by Teachers College address the Bureau of Publications, Teachers College, Columbia University, New York.

You will find the following books useful. Two or three of them should be in every teacher's library. McCall's book is rather advanced, but it is too good to leave out. Most of these books devote space to intelligence tests as well as educational tests.

Gilliland, A. R. and Jordan, R. H. *Educational Measurements and the Classroom Teacher.* New York: The Century Company, 1924. 269 pp.

Gregory, C. A. *Fundamentals of Educational Measurement.* New York: D. Appleton and Company, 1922. 382 pp.

McCall, W. A. *How to Measure in Education.* New York: The Macmillan Company, 1922. 416 pp.

Monroe, W. S. *Measuring the Results of Teaching.* Boston: Houghton Mifflin Company, 1918. 297 pp.

Pintner, R. *Intelligence Testing.* New York: Henry Holt and Company, 1923. 406 pp.

Pressey, S. L. and Pressey, L. C. *Introduction to the Use of Standard Tests.* Yonkers, New York: World Book Company, 1922. 263 pp.

Trabue, M. R. *Measuring Results in Education.* New York: American Book Company, 1924. 486 pp.

CHAPTER V

COMBINING INTELLIGENCE AND EDUCATIONAL TESTS

THE next step in better testing was taken when intelligence and subject-matter tests were given to the same child. The intelligence test told us what he *could* learn, the subject-matter test told us what he *had* learned. What was more natural than to bring these two facts together?

In any estimate of the success of teaching, a knowledge of the fundamental ability upon which successful learning depends is essential. It is astonishing how many supposedly important things in education turn out to be relatively immaterial. One investigator after another—I am thinking particularly of J. M. Rice, but there are others—has shown the relative unimportance of the course of study, the time devoted to a subject, the textbook used, the size of class, and the ability of the teacher. It is not so much that these things ought not to matter as it is that they are so handled that they do not matter. In this connection Courtis says that "the combined effect of all these is but one-half to one-tenth that of a single other factor—the individual differences in the native ability of children." He finds native ability to be "the great determining factor" and declares that school work is everywhere inefficient, because this factor is increased, not by getting better schoolhouses, equipment, and textbooks, nor by devoting more time to teaching and drill, but by making for each child an individual adjustment based on measured need.

If the native ability of the pupil is so essential and if efficient teaching depends upon knowledge of it and action in conformity with it, then any means at the disposal of the school for determining, even with a considerable margin of error, what this native ability is, should be made welcome and utilized for all it is worth. The intelligence tests purport to give us an approximation to a measure of the ability of every individual pupil.

The subject-matter test when unsupported by intelligence testing only tells us where the child is in reference to an average obtained from a large number of children. The author publishes an average for each grade or for each age. These averages are the *standards* for the test in question, and this is why it is called a standardized test. The standards are sometimes called norms. Thus the fifth-grade norm for the Burgess Reading Test enables us to compare with some satisfaction the work of a given fifth grade with the work of fifth grades in general.

Even here great injustice is often done to teachers, because some fifth grades are not so capable as the average fifth grade. When, however, we take individual pupils into account, the comparison of their scores with the grade norms does not help us so much as a comparison of their achievement with their intelligence. The teacher is concerned with individual pupils. This is fundamental. Surveys of general conditions are well enough, but are they not made in order that these conditions may be improved? And what other basis is there for improving general conditions than the improvement of Tom, and Dick, and Harry, of Mary, and Susan, and Jane?

The subject-matter test is insufficient. It has often worked grave injustice both to pupils and teachers. A child whose score is below the grade norm may have done well according to his lights—may have done far better than another who with greater powers failed to live up to them, yet secured a higher score.

A STANDARD FOR EVERY PUPIL

Evidently what we need is a standard for every pupil. Nature has made each child a separate entity. She has mixed the elements in him according to an individual formula and has given him possibilities which are essentially personal. A characteristic pattern has been sketched for him, and it is the task of the school to fill in the outlines and bring out the colors. The more the teacher knows about these abilities and aptitudes the more successful he will be. The teacher needs to know how fast and how accurately the pupil should be able to add, how rapidly and with what understanding he should be able to read, how hard the words are which he should be able to spell, and with what skill he should be able to use his mother tongue.

Now if, as I have indicated, the ability to learn these things is a matter of intelligence, is it not clear that we must seek the standard for each child in the intelligence which he possesses? Let us then give an intelligence test in connection with our test in arithmetic or reading or spelling. We can then compare his score in arithmetic with what his score in arithmetic should be according to his intelligence. We may likewise compare his reading ability with the reading ability that a boy of his learning capacity should have attained. Thus it comes about that instead of a general standard or grade average, we may have as many different standards as there are pupils.

At this point it will doubtless occur to you that a child may have a given degree of general intelligence without having a corresponding degree of spelling intelligence or arithmetic intelligence. This is indeed true. When a child learns arithmetic he brings to the task not only a certain general ability but also a particular ability which we may call an ability in arithmetic. Similarly, children possess special abilities in learning foreign languages, in manipulating tools, in music, and in fact in all the things they do, both great and small. Some time we may have tests for certain special abilities—perhaps for those corresponding to each of the school subjects. If that time comes, we shall then be able to supplement the subject-matter score and the general intelligence score with a third significant measure.

Meanwhile, however, we shall be foolish if we reject the measure of general intelligence, because it does not tell us all we want to know. These different special abilities are likely to vary with one another and with general intelligence. The child who has high ability to learn to read is likely also to have high ability to learn to spell or compose or solve problems. If a child cannot easily learn one subject, he is not likely easily to learn another subject. With a margin of error, therefore, which, while it is appreciable, does not destroy the significance of our findings, we may profitably use the present intelligence tests as part of our testing program. By so doing we may set up for each child a reasonably accurate standard valid for him and not merely for a group to which he may indeed belong, but with which as an individual he can by no possibility be identified.

While asserting the desirability of combining intelligence and subject-matter testing I wish to avoid giving the impression that I regard these measures as sufficient. No information about the child is unimportant, although it may be, and generally is, must less important than these two findings which the tests afford. Certain supplementary data are often required to prevent unwarranted conclusions and unjust actions. We shall want to know more concerning the child himself than our present tests permit us accurately to measure—his attitudes, his appreciations, his handicaps, and especially his moral qualities of persistence and effort. Under certain circumstances matters other than these may have important bearing. If the course of study to which a child has been subjected is markedly out of line, his failure to give a good account of himself may necessitate taking this fact into account. Similarly, in at least a few instances, wide differences in matters of equipment, textbooks, teaching ability, size of class, and time allotted to the various subjects will come in for consideration. The tests do not tell the whole story.

I have said before that the teacher will never be relieved by all the tests in the catalogues of the publishers from exercising professional judgment. Cyril Burt made a fine statement of this matter when he said: "Tests, infinitely scientific, can still be but the beginning, never the end, of the examination of the child. To take a young mind as it is, and delicately one by one to sound its notes and stops, to detect the smaller discords and appreciate the subtler harmonies, is more of an art than a science. The scientist may standardize the method; to apply that method and to appraise the results needs the tact, the experience, the imaginative insight of the teacher born and trained."

THE METHOD OF COMBINING INTELLIGENCE AND EDUCATIONAL TESTS

I have identified intelligence with ability to learn. A child's mental age may be said to be his efficient age with reference to school work. If our intelligence tests were entirely satisfactory, there is no question but that the mental age of a child should speedily replace his life age or chronological age as an estimate of his maturity for learning and instructional purposes. Even as things are, that is, with all the admitted shortcomings of the intelligence tests, these tests are so far superior to anything yet devised, that the mental age may well be regarded as the best measure of ability to learn.

When, therefore, we give an educational test and find out how much a child has learned in the subject in question, the bringing of this measure into comparison with the mental age shows an important fact. It shows whether or not the child has lived up to his expectancy.

Before we can make the necessary comparison, the intelligence and the achievement in the subject-matter test must be expressed in the same units. We cannot compare a mental age of ten with a rate score in reading of 155. That is, we cannot do this until we have converted the rate score of 155 into an age score. It is just as necessary to do this as it is to change one-half and one-third to a common denominator before we can add them.

I showed in the last chapter the value of the grade score —the practical utility of being able to say as directly as possible, "This is fifth-grade ability in spelling, that is seventh-grade ability in reading." But here we are confronted with the fact that we want not only to say such things as those but also such as the following: "This child's reading is so many units above the personal standard set by his intelligence," or "This pupil's performance in arithmetic is so many points below what his known intelligence would lead us to expect."

When, therefore, we bring educational and intelligence tests together, we must do so on the basis of age rather than grade—this for the very good reason that the intelligence-test scores are all expressed in mental ages rather than mental *grades*. At the same time we must recognize that when the test makers convert, as they must for this purpose, test scores into age scores, they commonly err in representing the smaller age scores as too small and the larger age scores as too large.

Test scores, then, both mental and educational, must be converted into age scores before they can be easily compared. We first find out the average score for unselected children of each school age. A score in rate of reading of 155 (Monroe) corresponds to an age of twelve and a half. Therefore, if a child scores 155 in rate of reading according to Monroe's test, we say his *reading age* (or his *achievement age* in reading) is twelve and a half. Suppose this child has a mental age of only eleven years. We may then say of him not only that he is doing as well in reading as the average child of twelve and a half years, but also that he is doing a year and a half better than his own standard of eleven years would lead us to expect. We may compare the eleven-year mentality and the twelve-and-one-half-year achievement either by subtraction or by division. If the achievement age (12½) is divided by the mental age (11), we obtain the *achievement quotient* (also called the *accomplishment quotient*). In the case of a child whose mental age is 11 and whose achievement age in reading is 12½, the achievement quotient in reading is 12.5 divided by 11, or 1.14. An achievement quotient is usually carried to two decimal places and then the decimal point is omitted. This particular quotient would therefore be written 114. This achievement quotient expresses the fact that in the subject in question—namely, reading rate—the child has reached a point of development 14 per cent greater than the average of pupils of his mental age.

If a child's achievement age in a given subject equals his mental age, his achievement quotient is 100. This expresses the fact that he is doing as well in that subject as his mental age leads us to expect and no better. Correspondingly, if his achievement age is less than his mental age, his achievement quotient is less than 100. In other words, a quotient of less than 100 indicates that in the subject or subjects to which it refers the pupil is not doing as good school work as the average pupil of his mental age.

Since, for the purpose of comparison with mentality, achievements in all subjects are expressed in terms of age—that is, in the same units—we may combine these separate achievement ages into a more general achievement age. For example, in the battery of tests known as the Illinois Examination, three kinds of test scores are converted into achievement ages. These are rate of reading, comprehension of reading, and arithmetic. If a child reads at the rate of 150 words a minute with a comprehension score of 11, and if he has an arithmetic score of 44, his achievement ages on these three items are 12 years, 11.5 years, and 10 years. These last figures may be averaged for a more general achievement age. This gives 11.2 as the achievement age in reading and arithmetic combined. Here equal weight is given to rate, comprehension, and arithmetic. Or, in other words, twice as much weight is given to reading as to arithmetic. If the teacher desires to give as much weight to the arithmetic as to the reading, doubling the achievement age in arithmetic will accomplish the result. In that event the average achievement age for the child referred to above would be 10.9 years.[1]

In sizing up the effect of the school upon a child the relation of achievement age to mental age is important. If the achievement age is greater than the mental age—that is, if the achievement quotient is greater than 100—factors other than intelligence must have been favorable. The school has doubtless done its work well. Of course, there are personal characteristics other than intelligence which may have been operative. The most important of these is effort. The child who tries hard is the one who makes a better showing than his mentality leads us to expect. But since unusual effort is quite as likely to be elicited through the superior work of the school as for any other reason, it is probably not inappropriate for the school—that is, for the teacher—to take credit for high achievement on the part of a pupil.

---

[1] If the reader is interested in pursuing this matter further I suggest that he read the manual which accompanies the Stanford Achievement Test and the manual accompanying the Illinois Examination. The fundamental memoir on the subject is Franzen's *The Accomplishment Ratio*, published by Teachers College, Columbia University.

By the same course of reasoning, however, the teacher must be prepared to assume the responsibility if the achievement age is less than the mental age. It is true that the case here is not always clear. Disturbing influences at the time the subject-matter test was given may lower the child's score. The point is that disturbing influences *always* reduce a score. There are no circumstances in virtue of which a child may secure a score above that of which he is capable. Again, a child may secure a low score in the educational test because of irregular attendance, over which the school may have had no control. But even when all possible allowance has been made, it is safe to say that if achievement age is less than mental age an investigation of the reason should be made. If investigation fails to reveal clearly a contributing cause beyond the control of the school, then the school must assume the responsibility for the condition.

I plead, therefore, for a combination of intelligence testing and subject-matter testing. We need this combination in order to secure a personal standard of work for each child. Moreover, without the intelligence tests the subject-matter tests may work great injustice to the school and to the teacher.

The story is told of two first-grade teachers in the same city. One was securing excellent results, the other the reverse. The latter worked hard but apparently to little purpose. Her children read haltingly, wrote poorly, and were uncertain in their number work. She was discouraged, and the superintendent had made up his mind to get rid of her. An intelligence test, however, revealed the fact that the average mental age of her pupils was not only below the normal age for the grade but was actually quite two years less than that of the pupils taught by her more fortunate colleague. Yet she was being expected to produce the same results—a task altogether beyond human power. This teacher was in reality a better teacher than the one who apparently secured better results.

On the other hand, the use of intelligence tests without subject-matter tests is one-sided. No doubt a child whose mentality is six or six and a half can safely be placed in the first grade. There is also little doubt but that a child of seven and one-half (mentally) can do the work of the second grade if he has received proper instruction. He certainly cannot do so if he has had no instruction or very poor instruction. In general, the evidence seems to be that when pupils evince a mentality corresponding to a given grade they can do the work of that grade if the school has done its part. It is precisely the function of the educational tests to tell whether or not the school *has* done its part.

The real truth about a pupil or a class, then, is only made evident when to a knowledge of mentality is added a knowledge of achievement. If mental age may be taken as *potential* ability, achievement age may be taken as *effective* ability. The one shows a child's chance to learn, the other what he has learned; the one measures his ability as a possession, the other his ability as put to use. Since achievement age, however, indicates ability functioning, it affords an important supplement to and commentary upon mental age. Moreover, it affords, when related to mental age, a stimulus to effort or a basis of satisfaction according as it is less or more than the mental age.

CHAPTER VI

NEW-TYPE EXAMINATIONS

THE conviction that the rating of pupils ought to be made very much more accurate has led not only to the use of standardized tests but also to the adoption of entirely new examination methods on the part of up-to-date teachers. It is not too much to say that the results of the so-called "test movement" have been as pronounced in their effect upon local testing as they have in any other direction. The makers of intelligence tests and of standardized subject-matter tests early developed certain techniques which had not previously been known. Although each of these tests had to cover a wide field, it had to do so without taking much of the pupil's time. Moreover, it was necessary—in view of the large number of pupils and the important issues involved—that the answer papers should be scored quickly and reliably. These are important advantages in connection with any test, no matter what may be its origin or purpose. Teachers have, therefore, been quick to note that the techniques by which these advantages were secured had no essential reference to standardized material—that, in other words, they were applicable to tests for special, immediate, and local needs.

Moreover, the teachers who were receiving and using the products of research had become convinced that their rating of examination papers was faulty. They believed, however, that this faulty rating arose not so much from their own fallibility as from the nature of the tests with which they were dealing. They realized—or at least a few of them did—that there was no way in which the discussion or essay type of examination could be rated so that two or more competent teachers would agree on the rating of a given paper. Under these circumstances these teachers naturally turned to the improvement of tests as a means of improving their ratings of pupils.

Accordingly, certain new methods of conducting class examinations have come into use. For want of a better name these methods are called "new" and the examinations which embody these methods are called "new-type examinations." I should prefer to call them objective examinations, for a large part of the purpose in employing these so-called new techniques is to make the appraisal of the pupil's response so independent of the teacher's personal judgment that anyone who knows the correct response can rate the paper. When this condition is brought about we have an "objective" condition. To call a thing new is merely to declare it to be recent. Sooner or later such a name is going to become absurd. One may rummage in an old attic and come upon "new" spellers, "new" geographies, and "new" readers. If subsequent techniques of test-making are developed, what name is left for us to give to the ones which we now call new? But although the term new-type as applied to the present techniques is an inferior term, the sanction of usage is with it, and I shall not attempt to introduce a change.

I have said that the essential feature of the new-type examinations is their objectivity. In order that they may have this feature, the questions of which they are composed are special, narrow, and definite. In this they are opposed to the general, broad, and indefinite questions of the "old-type" examination. The old-type examinations are the ones in which the examinee is asked to "define and illustrate," to "discuss," and to perform similar tasks involving the writing of an essay in response to each question.

Here again there is a distinction between the old- and new-type examinations. In the former the answer is long, discursive, and impossible to rate accurately. In the latter the answer is short and unequivocal. It is often a word or a number or a line drawn around, under, or through an expression. In saying that the answer is unequivocal, I mean, in the first place, that it is either wholly right or wholly wrong. No part credits are allowable for different degrees of rightness, because no degrees of rightness are recognized. If I ask the question, In what state is Pikes Peak? the answer "Colorado" is wholly correct. Any other answer is wholly wrong. No more credit is given to the child who locates Pikes Peak in Wyoming than to the child who places it in New Hampshire. In the second place, the unequivocal character of the answer in the new-type examinations is indicated by the fact that everyone who is well informed concerning the matter in hand agrees as to the correct answer. No one, for example, who knows the location of Pikes Peak will admit for a moment the correctness of any other answer than Colorado.

Broadly speaking, there are two kinds of new-type examinations—namely, the *recall* examination and the *recognition* examination. In a recall examination the person who is being examined is required to originate the answer. In the question about the location of Pikes Peak the examinee, if he answers correctly, must originate the response "Colorado." In other words, he must recall this answer as a proper response to the question. Accordingly, an examination which makes this kind of requirement is a recall examination. On the other hand, in a recognition examination the correct answer is actually given as part of the examination. The examinee is required, however, to *recognize it as correct*. For example, I may get at the idea that Pikes Peak is in Colorado not by asking the pupil to name the state but by telling him to underline the word which makes the following sentence true:

PIKES PEAK IS IN—
Vermont, Wyoming, Colorado, California, Virginia

In the recognition examination a great many devices are used to prevent the recognition of the correct answer by anyone who is not well grounded in the matters with which the question deals.

THE RECALL EXAMINATION

As I have said, this kind of examination involves questions in which the examinee originates the response. Recall questions have certain decided advantages over recognition questions. Teachers rightly feel that an examination in which the pupil must originate the answer is considerably more exacting than one in which he merely selects the correct answer from several possible answers.

Yet the recall type of question presents many difficulties both as to construction and rating. In the first place, as its name suggests, it is addressed primarily to the memory. Some makers of standardized tests have tried desperately to construct thought questions of the recall type. They have been successful only to a limited degree.

In the second place, the unequivocal character of the answer—desirable if the test is going to be objective—is difficult to maintain. Consider the question, Who wrote *Paradise Lost?* We say that the correct answer is "Milton." But suppose the child answers "John Milton." At once you say this is correct too. But it is not the same answer and so there are two correct answers. Then there is the bothersome question of answers which have a certain formal correctness but which are quite unsatisfactory on other grounds. Suppose a child responds to the above question with the answer "a man." This is true enough but it is inconsequential. I think any teacher would rate such a reply as wrong. Yet it is certainly true even though it is not satisfactory. Perhaps some timid souls would give the child "part credit" for this answer. Let us make another supposition. Let us say that a child's answer to the question was "a blind man." Is this any more correct than the response "a man"? Would you think it entitled to part credit? Finally, suppose a child really knew a lot about Milton and his writings, but for some reason could not recall his name. Everybody has such lapses and they are quite understandable. Let us suppose that this child answers the question, Who wrote *Paradise Lost?* by writing, "The man who was Latin secretary under the Commonwealth and who wrote the *Areopagitica*." Such a child has some real knowledge of Milton, but according to the usual acceptance he has not answered the question correctly.

There are two forms of the recall examination, namely, (1) the question-answer form and (2) the completion form.

Question-answer. According to this form the requirements—that is, the items of the test—are set up as questions.

> Who wrote *Paradise Lost?*
> In what year was Washington inaugurated?
> What state is next north of Maryland?
> What is the name of a four-sided plane figure?

Each of these questions has a definite and fairly unequivocal answer. The extent to which this fails to be the case and the consequent failure in perfect objectivity is often compensated for by the fact that the pupil must rely upon his own unaided efforts. The justification for this is the fact that life presents many situations in which the answer to a personal problem is of the recall type. We need to remember when a thing happened, what this person's name is, and where we take the left-hand turn. Although, therefore, as we shall shortly see, examination questions of many varieties have been constructed on the recognition plan, there is still a place for the recall examination of the question-answer form.

Completion. The completion form of the recall type is richer in its possibilities than the question-answer form. If instead of asking, Who wrote *Paradise Lost?* I represent the requirement in the form of a statement with the essential word omitted, I am using the simplest case of the completion form. For example:

> *Paradise Lost* was written by ————?

The resources of the completion form, however, are best seen when more than one blank appears in a given connection:

> A river with all its ———— is called a river ————.
> ("Tributaries" and "system" to be supplied.)

An excellent use of the completion examination is made in the Stanford Achievement Test. In one portion of this test paragraphs are presented for the pupil to read. A part of the material is left out, and he is called upon to insert it. He is told to write just one word in each blank. With this limitation on the filling of these blanks I think you will find they are almost, if not quite, inevitable. For example:

> Once there lived on a mountain near a village an immense giant whose cruelty kept the people of the village in great terror. However, there was one person in the village who was not afraid of the giant. This was a young soldier who carried a magic sword that a fairy had given him. Once when the ———— came down from the ———— the soldier attacked him with his magic ———— and killed him.

Again:

> It is well established that the bee, which is commonly supposed to be so industrious, really works only two or three hours a day. The man who works eight or ten hours a day is therefore far more ———— than the ——.

The completion technique has been widely and usefully employed, though rather less frequently of late years. The remarkable development of various forms of recognition tests has overshadowed the completion test. It has great advantages, but it is not easy to score, especially in its more elaborate forms. The trouble is that the answer is not unequivocal. In this respect a completion question involving the filling of one blank suffers substantially the same limitation as does the question-answer form. Quite commonly, however, a completion question calls for the filling of more than one blank. Under these circumstances it is even possible to set up indefinite conditions—that is, conditions which permit of numerous entirely different answers. The question then becomes valueless. For example:

> ———— was the leader of the ——— army.

The person who made this completion exercise might have intended it when completed to read: Washington was the leader of the Continental Army. But a pupil might complete the sentence in any one of a number of different ways, such as: Lee was the leader of the Confederate Army.

More frequently the completion exercise involving more than one blank merely permits words of slightly different meaning to be used in the blanks. This, however, multiplies the number of combinations and each combination constitutes a different response to the entire exercise. Several of the combinations may have a meaning so close to that of the accepted answer that full credit will seem to be justified. Other combinations will appear to be entitled to some credit but not to full credit. Still other combinations will be thought to be valueless. Where to draw the line is the main difficulty, and here is where one person will differ from another in rating the same paper.

I am not dwelling on these points to raise in your minds a doubt of the value of either the completion or the question-answer form of the recall type of examination. I have a decided belief in their value, especially when they are combined with certain of the recognition types. I think the merits of the recall examination have not been sufficiently exploited, especially in the direction of devising thought-provoking questions. Moreover, I feel that the exaction in virtue of which the examinee must originate the response rather than recognize it is one which we cannot afford to abandon.

THE RECOGNITION EXAMINATION

A remarkable number of recognition techniques have been developed. The makers of intelligence and subject-matter tests have shown surprising ingenuity in the methods they have employed. Most of these methods are useful for the teacher. I think you will experience real satisfaction if you will adopt some of them in your testing. I have distinguished four major techniques under the head of the recognition examination. These are (1) the alternative response, (2) the multiple choice, (3) matching, and (4) combinations of these.

Alternative response. As the name indicates, the alternative-response test is one in which the pupil is called upon to make a choice between two, and only two, possible answers.

Life—and I think all these matters must be tested by their conformity to life—presents endless situations where choice must be made between just two alternatives. Whether to take the right fork of the road or the left, whether to work or rest, whether to speak or be silent, whether this proposal is sound or unsound, whether that statement is true or false —in endless variety these and hundreds of other alternatives call for choice. Most of us are more competent in some fields than in others. In the fields of our competence we may be expected to make a high percentage of right judgments. Indeed, the success we have in making these judgments may be taken as a measure of our competence. Accordingly, a test of competence may be constructed in which a series of opportunities for choice between alternatives is offered.

The true-false test. One way in which this opportunity for choice may be offered is by presenting a series of statements of which some are true and others false and by requiring the pupil to indicate which are true and which are false. A test of this sort is called a true-false test. For example, the statement may be made, "The regular coast line of Africa has hindered its development." The pupil may then be asked to show whether this statement is true or false. If he says it is true, he asserts in effect that Africa has a regular coast line, that its development has been retarded, and that the regular coast line has hindered the development. If he says it is not true, then he denies one or more of these facts. Since competent people call the coast line of Africa regular in comparison with that of other continents, and since it is similarly agreed that the development of Africa has been "hindered," and finally since a causal relation between these two facts is generally believed to exist, the pupil who thinks this statement is *false* betrays a lack of competence in respect to these matters.

A true-false test may consist of fifty or a hundred or even two hundred statements. In fact, in order to be a satisfactory test the number of statements must be large. About half of them should be true and the other half false. Sometimes the pupil is directed to indicate his decision by making a plus sign after a statement which he thinks is true and a minus sign after one which he thinks is false. Sometimes the words "true" and "false" are mimeographed or printed after each statement. The pupil is then directed merely to underline or draw a circle around the word which applies to the statement in question. Still another way of having the pupil indicate the truth or falsity of a statement is to print the letters T and F in front of it. The pupil is then required to underline or to encircle T if the statement is true and F if it is false.

The following ten statements form part of a true-false test on Monroe's administration. The pupil is to underline the word *true* if he regards the statement as true, and the word *false* if he regards the statement as false.

1. The Missouri Compromise provided that Missouri should be admitted to the Union as a slave state. True False
2. The first passenger railway in the United States was built during Monroe's administration. True False
3. John C. Calhoun was the chief supporter of the Missouri Compromise. True False

4. The National Road to the West began at Cumberland, Maryland. True False
5. The Monroe Doctrine was agreed to by the European countries in 1823. True False
6. Lafayette visited the United States during Monroe's term of office. True False
7. Monroe was the fifth president of the United States. True False
8. Monroe was president for only one term. True False
9. Just before the Missouri Compromise the number of slave and free states was the same. True False
10. Monroe was elected president only after a bitter contest True False

A variation of the true-false test occurs when the examiner, instead of presenting to the pupil a series of statements to be recorded as true or false, presents a series of questions to be answered by "yes" or "no." Instead of the statement, "The Civil War was fought to destroy slavery," or the statement, "The spoils system was brought into prominence in Jackson's administration," we may have the questions:

Was the Civil War fought to destroy slavery? Yes No
Was the Spoils system brought into prominent use in Jackson's administration? Yes No

Other kinds of alternative-response tests. There are, of course, many other ways of setting up alternative response tests, and some of them have had rather general use. One of the commonest is the so-called "same-opposites" test. This is an examination of the pupil's knowledge of English words. It has been extensively used in intelligence testing. Pairs of words are printed together and the pupil is required to show whether the words mean the same or opposite. For example, the following pairs may occur, the pupil being required to underline "same" if the two words of a pair mean the same or nearly the same, or the word "opposite" if they have opposite or nearly opposite meanings:

new—old	same	opposite
broad—wide	same	opposite
liberty—freedom	same	opposite
calm—stormy	same	opposite

When English usage is tested, it so happens that most constructions have one very common error in addition to one right form. This suggests an alternative response test, and this kind of test has been frequently used to measure a pupil's knowledge of particular English usages. For example, a test may contain items like these, the pupil being required to select the right word in the parentheses:

I feel (bad, badly).
He (don't, doesn't) like spinach.
(Who, Whom) did you meet?

Still another of the many uses of the alternative response idea arises in spelling, where again there is frequently one very common misspelling of a word. The following is taken from the Purdue English Test. The pupil is to underline "right" or "wrong," according as he believes the word to be correctly or incorrectly spelled.

beginning	right, wrong
privilege	right, wrong
indefinate	right, wrong
digestable	right, wrong
disappear	right, wrong

A great deal has been written about the merits of the alternative-response test. It has had ardent supporters and vigorous enemies. A method of scoring is used to discount the influence of guessing. The method is to deduct the wrong answers from the right answers for the pupil's score. This deduction is made, because on items of which the pupil is ignorant he will, on the average, guess right half the time. If he has failed on a certain number of items, there are likely to be, according to this theory, an equal number of items on which he has succeeded without really knowing the answer. Therefore, to find the number of items he really knows we must deduct from the right answers a number equal to the number of wrong answers. This formula will give a negative score whenever the pupil has more items wrong than right. Negative scores are generally regarded as zero.

The weight of evidence seems to justify the alternative-response test if it is carefully made and contains a large number of items. Such a test has an advantage over the more usual discussion test, for it permits the teacher to cover a great many topics with a small expenditure of time on the part of the pupil. If the items are not too long, elementary-school pupils may answer as many as fifty in ten minutes. The test may, therefore, easily become representative of the entire field of instruction to which it refers. The common practice of a student in attempting to guess what the teacher will ask on the coming test is rendered futile because the teacher who uses an alternative-response test is likely to ask for everything of importance.

It is also claimed for these tests that, due to their wider range and more objective character, they are fairer to the pupil. The point is made by several investigators that when pupils have had an opportunity to express their preference for the traditional or the alternative response test, they have voted strongly for the latter.

Most of the arguments, however, for the alternative-response test are likewise arguments for all kinds of new-type examinations. Moreover, some of the disadvantages—especially those due to the workings of chance—apply in a less degree to the type I am about to suggest. On the whole, therefore, it appears that the alternative-response test has decided value, but that it should be used as only part of the test techniques which the modern teacher has at his disposal.

Multiple choice. The main trouble with the alternative-response test arises from the fact that guessing plays so important a part. If the pupil merely has two choices, he stands an even chance of being right even if he is quite ignorant of the matter in hand. Accordingly, the type of test has been developed in which the pupil has to make one or more choices among *several* items. The likelihood of being right by mere chance is thus greatly reduced. For example, in the Purdue English Test, under the head of "Literary Information," a series of sentences is given each containing in parentheses four names or terms. The pupil is to "underline the one which is needed to make the sentence a true statement." Here are a few of the sentences:

> The character Brutus appears in (Hamlet, Macbeth, Julius Cæsar, Merchant of Venice).

"What is so rare as a day in June?" is a quotation from (Lady of the Lake, The Vision of Sir Launfal, Evangeline, Ancient Mariner).

A novel by (Benjamin Franklin, Helen Hunt Jackson, Hawthorne, Harriet Beecher Stowe) was written in defense of the American Indian.

Unlike the alternative-response type of test, the multiple-choice type has great range and variety. If all the other new-type techniques were abandoned, this one would still furnish a reasonable basis of testing. It can, of course, be used to test information, but it also lends itself to the testing of thought. From its very nature it requires *selection,* and selection implies thought.

The multiple-choice technique may, and often does, require the examinee to infer from specific instances the general principle according to which the subsequent selection is to take place. This sometimes involves *classification* and sometimes the apprehension of a relationship. For example, a child may be told to select from the second set of the following words the one which belongs to the first set:

First set:
  Peach, plum, cherry
Second set:
  Sweet, red, pit, seed, pear, tree

The likeness—the common characteristic—among the things named in the first set must first be perceived. In other words, these things must be *classified* as fruits. Then this idea must control the selection of the right word from the second set.

Again the multiple-choice technique may set up the ideas (1) *president* and (2) *republic.* It may then give a third idea, *king*. It may next require the examinee to select from a series of words the one which sustains the same relation to *king* that *republic* does to *president*. The series of words may be: *queen, nobility, senate, empire, kingdom*. Here the pupil must see the relationship between *president* and *republic* as the relation of a ruler to the state he rules. He must then find in a series of words the name of the state which a king rules. To keep these things in mind and to resist the suggestion of *queen* as related to *king*, of *nobility*, and of *empire* as states which may be ruled—but not according to the specifications—to do these things requires straight thinking.

Moreover, the mutiple-choice technique is not limited to the selection of *one* item from a number. It may call for the selection of several. It may even impose an additional condition, such as the arrangement of the items selected according to some required plan. For example, in a history test the pupil may be required to "select from the following events the three that had bearing on nullification and arrange them in the order in which they happened."

It is true that the multiple-choice techniques that require thought have been used, for the most part, in intelligence tests; but there is no reason why this should remain true. We are agreed, are we not, that we should teach children to *think* in mathematics, geography, history, and science? But if we do not in our testing require them to think, if we do not measure their ability to think, and if we do not make school success depend upon this measured ability to think, our belief that thinking should be taught will remain ineffective.

One of the things, therefore, which you as teachers can do in a research way is carefully to devise new-type examinations, especially those which require pupils to think. Suppose, for example, that a body of teachers, such as the American Historical Association, was to devote a year to this matter in the field of United States history. There is already in existence a rather well-understood series of objectives in teaching history. Pupils must first know the facts of history. There is no dispute about that. Then they should be able to see relationships of many sorts, to appraise events and men, to appreciate meanings, and to weigh evidence. The members of the American Historical Association would add enormously to the value of the teaching of history if they were to devote themselves to devising tests of all these abilities—tests according to modern objective techniques, tests which covered each topic and period of United States history as fully as the public schools may be expected to cover it. Such a body of test material put into book form and made available to teachers *and pupils* would give to history teaching a scope, dignity, variety, and precision which it now lacks.

Not the least of the advantages obtained when the teachers of Latin recently put through the Classical Investigation was the test material which they developed in connection with the project. One of the things which the investigation showed was the lack of anything like agreement among the teachers as to what they were trying to do. Part of the reason for their failure to see their objectives more clearly was the absence of any convincing way of finding out the extent to which any objective they might entertain was being reached. With the large addition to the resources for testing Latin abilities it has become possible to measure success along the lines of many of these objectives. Moreover, through these means it is not too much to expect that the objectives will be more clearly conceived and more generally agreed upon.

The information question. The multiple-choice technique is perhaps most frequently, though not always most wisely, employed in making up information questions. For example: [1]

(   ) What state mines the most soft coal?
1 2 3 4
Wisconsin Illinois Vermont Utah

The child is expected to write the number of the state which mines the most soft coal in the parentheses in front of the question. Again:

(   ) Which seaport is nearest Japan?
1 2 3 4
Seattle New Orleans Galveston Baltimore

(   ) To which country does the United States export the most petroleum?
1 2 3 4
Rumania Mexico Holland Russia

The multiple-choice type of examination has been so commonly employed in connection with information questions as to create in the minds of some people the impression that this technique does not lend itself to the testing of thought processes. Nothing could be further from the truth. The multiple-choice technique has many uses; and some of these uses, as we shall shortly see, bring useful types of thinking into action.

---

[1] Buckingham, B. R. and Stevenson, P. R. *Geography Test, United States—Information—Problems.*

The thought question. Barr has shown several ways in which questions involving multiple choice may be used in history. Consider the following:

> Put a cross (X) before three of the following events which occurred about the same time (same general period). It is not necessary that these events should have taken place at exactly the same date.
>
> (a) The purchase of Florida
> (b) The Emancipation Proclamation
> (c) The settlement of Plymouth
> (d) The Declaration of Independence
> (e) The building of the Panama Canal
> (f) The battle of Vicksburg
> (g) The building of the Erie Canal
> (h) The Spanish-American War
> (i) The reëlection of Lincoln
> (j) The discovery of America [2]

Again,

> Put a cross (X) before the fact in the following list which was of the greatest importance in the nullification controversy:
> (a) The theory of nullification recalled the Virginia and Kentucky Resolutions in the assertion of the right of a state to declare an act of the national legislature unconstitutional.
> (b) The doctrine of nullification was set forth by Calhoun.
> (c) According to this doctrine each state was at liberty to nullify the law in case of dangerous exercise of power by Congress.
> (d) The theory of nullification arose out of the discussion of the tariff question.
> (e) Calhoun, as a strict constructionist, was an advocate of nullification.

---

[2] Barr, A. S. *Diagnostic Tests in American History.*

Note that in the first of the above questions from Barr, the pupil is called upon to choose more than one element from a group of elements and that all of the elements of choice must meet a single condition. This is one of the ways in which the multiple-choice type may be varied not only under this heading of "the thought question" but also under some of the other headings which follow. Whenever this is done the demands for thought on the part of the pupil are likely to be increased. The conditions controlling the selection generally become more difficult to apply when more than one thing is to be selected.

Definitions. It has always been hard to determine whether a child knew the meaning of words. Some interesting investigations have been carried forward in this connection. Children have been asked to check each of the words in a list whose meaning they were *sure* they knew, to put a cross in front of those which they *thought* they knew, and to indicate with a circle those which they were certain that they did not know. Again, children have been asked to write the definitions of words. Another method has been to require them to use the words in sentences. None of these methods has proved to be satisfactory. Yet some way of testing word knowledge is important. It is rather generally admitted that a knowledge of the vocabulary peculiar to a subject is one of the evidences of familiarity with the subject.

The multiple-choice technique gives satisfactory results. Instead of asking the child to tell whether he knows the meaning of "octagon," and instead of asking him to define it or to use it in a sentence, we may present the idea or definition followed by the name together with other names from which it must be selected as the correct one. For example: "An eight-sided figure is called a *Trapezium, Scholium, Parallelogram, Octagon.*"[3] Again the thing to be defined may be given in a sentence in which the key word must be chosen from a list of suggested words. For example: A *peck* is a fourth of a *Barrel, Bushel, Gallon, Keg.*[4] Or again: Anything *insured* is *Burned, Protected, Ruined, Increased.*[5]

Synonyms. This use of the multiple-choice method of examining resembles the one I have just described. The pupil is required to select from four or five words the one which is similar in meaning to a given word. A test of synonyms has special value for the teacher of English. This is one way in which a test may be given:

Allege (expand, cancel, accept, assert)
Tacit (noisy, open, silent, true)
Fallacy (truth, honesty, price, mistake)
Acumen (vanity, view, sagacity, cheapness)[6]

Opposites. The giving of opposites is scarcely recognized as a school exercise. It has, however, been used by the psychologists for many years in the rating of intelligence. It also has value for the teacher of English as a test of word mastery. A test of opposites may be constructed by the multiple-choice method. The first word being given, the opposite of it is to be selected from a group of words. For example:

---

[3] Terman, L. M., *Group Test of Mental Ability.*
[4] *Ibid.*
[5] *Illinois Examination.*
[6] Brandenburg, G. C. *The Purdue English Test.*

Difficult (hard, quick, soft, easy, common)
Always (sometimes, often, occasionally, seldom)
Unless (and, therefore, however, also, if)[7]

Cross-out. An ingenuous application of the multiple-choice idea is one which requires the pupil to select not the right but the *wrong* element in a situation. Having selected it, he very appropriately crosses it out. For example, in the following series of numbers the pupil is to notice the way the series is built up and to cross out the number that does not belong in the series:

2 4 8 10 16 32[8]

More evidently applicable to a school subject is the use of this cross-out technique in primary reading. For example, a child's understanding of sentences is assumed to be tested by having him cross out the extra word in such sentences as the following:

(1) I ball see a boy
(2) See dog my pretty book
(3) I like to gun play ball
(4) I cat can see my doll
(5) The baby is my book little sister[9]

In these exercises it is held—and I think correctly—that the pupil must have read the sentence with some understanding before he could select the intruding word. The same technique is used to test the recognition of words on the part of first- and second-grade pupils. In the following exercises the pupil is to cross out the thing that is not a "real word."

---

[7] Otis, Arthur S. *Group Intelligence Scale, Advanced Examination, Form A.*

[8] *Illinois Examination.*

[9] Pressey, L. W. *Second-Grade Attainment Scale, Form B.*

(1) See go uf be up
(2) Cat nid house man we
(3) Said wat that was not
(4) Down on am gen their
(5) Hen eat children going nent[10]

The possibilities of the cross-out idea are numerous. A child may be called upon to note the similarities among pictures or diagrams and to cross out the one which does not belong to the general classification. Some real thinking may be involved in meeting these requirements.

Similarities. This last idea, namely that of seeing the general classification to which a series of items belongs and crossing out the one intruder, requires much the same type of thinking as does the so-called similarities test. So far as I know the similarities test has not been used in examining pupils in school subjects, but there is no reason why it should not be so used. According to the variation employed by Otis, it consists in giving, first, a group of words which are alike in some particular and, second, a miscellaneous group of words containing one word, and no more, which is like those of the first group. The task is to see the sense in which the first words are alike and to select from the miscellaneous list the one word which has the same likeness. For example:

hat, collar, glove........hand, cane, head, shoe, house

In this instance the pupil is supposed to discern that *hat, collar,* and *glove* are alike in that they are articles of clothing and to select *shoe* from the next five words because it is likewise an article of clothing. Again it is possible to require a high degree of thinking. The sense in which the words first presented are alike may be far from obvious. To apprehend this likeness may require close observation and reflection. Moreover, the selection of the word which has the same characteristics but which must be identified among other words possessing certain other characteristics in common with it may be very far from easy.

---

[10] *Ibid.*

Analogies. The psychologists have long used analogies tests and have found them effective in measuring intelligence. Relationships must be discerned and applied. Originally the analogies test was in the form of a proportion. It did not use "multiple choice." For example:

hat:head::glove: ————

Here the correct answer is "hand," and the reading of the proportion is: "As *hat* is to *head,* so is *glove* to *hand.*" In this form of presenting analogies—a form in which the person examined originates the answer—there may often be more than one correct answer. It is precisely to obviate this difficulty that the multiple-choice technique has been employed. For example:

hand: arm: foot: (?) ......leg, toe, finger, wrist, elbow
coal: locomotive: (?): automobile......motorcycle, smoke wheels, gasoline, horn [11]

Like the similarities test, the analogies test has been used for the most part by psychologists. It should also be used by teachers. It employs an important type of thinking, and that type of thinking is especially useful in teaching and learning. How often do we as teachers employ analogies in our classroom teaching? Of course, the logician points out the pitfalls of "reasoning by analogy." But rightly employed the analogue is a powerful instrument in instruction. Not less important as indicating the extent to which a pupil understands relationships in a given subject, is the expression of those relationships through analogies. I suggest, therefore, the possibilities of the analogies test in your school work.

---

[11] Otis, Arthur S. *Group Intelligence Scale, Advanced Examination, Form A.*

Best reasons. In all our teaching we aim to have pupils understand the reasons for the facts they learn. Sometimes, it is true, we teach them reasons as facts; and always there is a large factual element in any true reason or cause. Nevertheless, even when "five causes" of the War of 1812 have been learned as such, there is some thinking in their recall and selection from competing facts.

In one of the geography tests the pupil is directed to "write in the parentheses the number of the correct or best answer." Such questions as the following are then proposed:

( ) Why should forests in the United States be conserved?
1. They increase the supply of oxygen
2. They make excellent places for hunting game
3. They supply building materials and decrease floods
4. They enrich the soil and shade the ground

( ) Why is New York State opposing construction of the proposed St. Lawrence waterway?
1. The St. Lawrence waterway would be very expensive
2. The proposed waterway could not be used in winter

3. New York would lose some of its foreign commerce
4. New York harbors would have to handle too much freight [12]

Here is another best-reason question, although it is not in the form of a question. It might have been taken from a physics paper. I have selected it, however, from the Terman Group Test of Mental Ability. The pupil is directed to put a cross before the best answer.

Freezing water bursts pipes because
1. Cold makes the pipes weak
2. Water expands when it freezes
3. The ice stops the flow of water

Pictures. In all that I have said above about multiple-choice tests I have referred only to material written or printed in words. Pictures may often be used. The pictures may be simple diagrams. The cross-out method may be applied to diagrams, so also may the similarities method, but the information question likewise lends itself to pictorial representation in a way which the teacher may find most suggestive.

About two years ago, I was asked to help the Department of Agricultural Engineering at Ohio State University in its testing. An examination had recently been given which had proved entirely unsatisfactory. It had consisted of five questions occupying perhaps a third of a typewritten page. These questions were to be answered in an hour. The chief difficulty with the test was that it was too hard and *too long*. Scarcely anybody finished it. The new-type test which we made up occupied ten of the same sized pages, yet every student finished it within the hour.

---

[12] Buckingham, B. R. and Stevenson, P. R. *Geography Test, United States—Information—Problems.*

One of the five questions on the unsatisfactory examination was: "Describe completely and accurately a moment arm." In order to answer this question a student would have to make a number of distinctions and either set them down or reject them in favor of a formulation which had general validity. This is a test—as the essay type of test usually is—of ability in English composition. Moreover, after the student had written out his answer, the instructor had to decide whether he had described a moment arm "completely" and "accurately." This was a difficult decision to make. He had to decide, too, what credit, if any, should be given for answers exhibiting various degrees of accuracy or completeness.

The question which took the place of this one about a moment arm was:

> "In the following diagram of a bell crank [See Figure 6 on next page], which distance, M, N, O, P, or Q, represents the moment arm of the force F? (Underline the correct letter in the answer square.)"

The correct answer to this question is O, the perpendicular distance from the point of support to the line of force. The student who underlined O in the answer square had answered "completely" and "accurately." The student who underlined anything else was entirely wrong. The student needed only time in which to record his answer. On the other hand, the instructor was able to rate the response almost instantaneously and with entire justice to the student as well as to the facts in the case.

Some testing lends itself particularly well to the use of drawings. For example, Figure 7 shows a system of pulleys. The formula $P = \frac{W}{N}$ being assumed, we may ask for the value of N or we may give W a value and require the computation of P. In either event we may give several different values in the answer square with the requirement that the correct one be underlined. The question to which Figure 7 belongs is: "Is the value of N 5, 6, 7, 8, or 9?" The student is to underline the correct number in the answer square. The point is that without resort to a diagram of the system of pulleys we "give the case away" in our description.

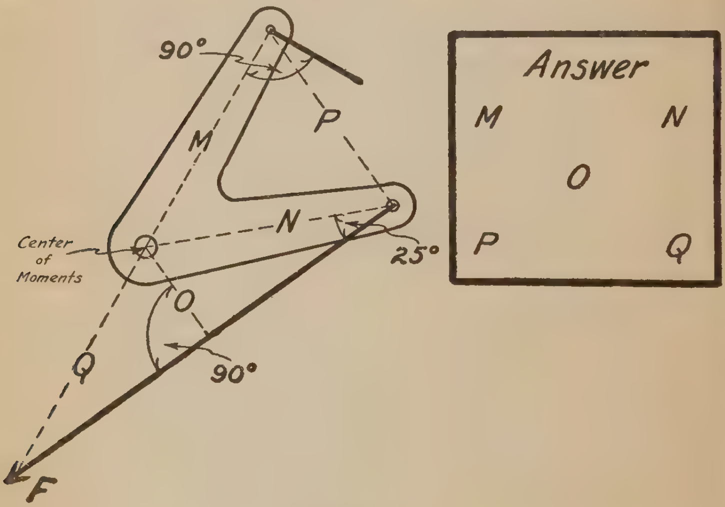

FIGURE 6.

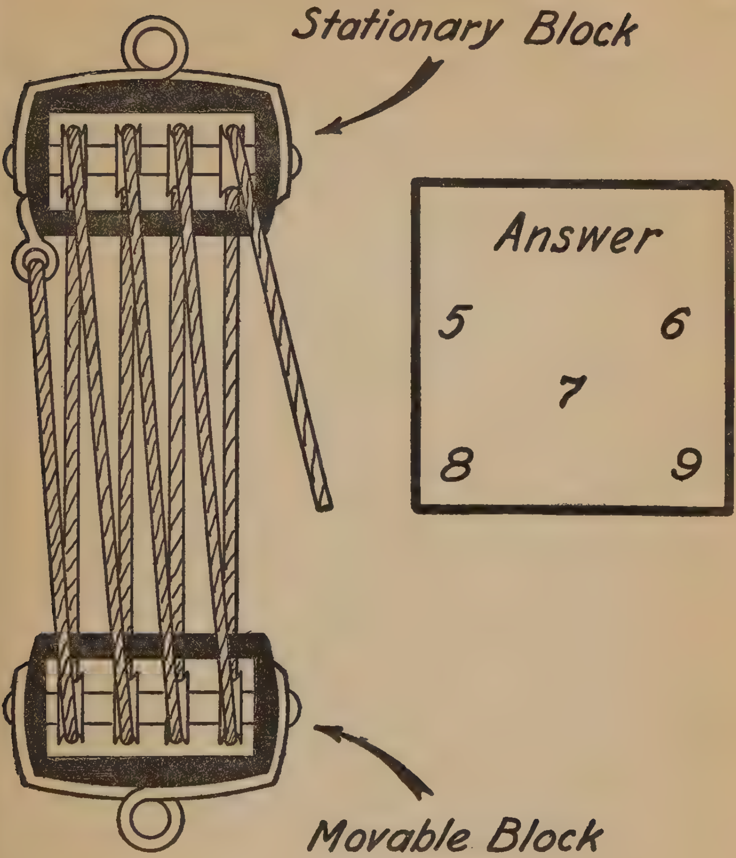

FIGURE 7.

Matching. One of the new types of questions requires the pupil to put together things that belong together. He must pick an item from one list and relate it to an item selected from another list. Thus the test is a sort of double multiple choice, with the added requirement that the two items selected shall stand in a defined relationship to each other.

Definitions. We have already seen that a pupil's knowledge of the technical terms of a subject may be tested by the multiple-choice method. It may also be tested by the method of matching. For example, in the definitions test in geography by Buckingham and Stevenson, the children are required to read each incomplete definition and to write in front of it the number of the word in the first column which properly completes it. To assist the pupil in finding the word in the first column, these words are put in alphabetical order. The requirement may be illustrated as follows:

COLUMN I	DEFINITIONS
1. coast 2. ~~moon~~	(4) Water falling from the clouds is called
3. ~~mountain~~ 4. rain 5. ~~snow~~	(6) The round shining body around which the earth revolves is called the
6. sun	(1) The margin or boundary of the land next to the sea is called the

Observe that there are more words from which to select than are needed. *Moon* is not needed; neither is *mountain* nor *snow*. If extra words are not included, the pupil can first select those whose meaning he knows. Then if there are only two or three left, he can use these correctly, provided he has any sort of idea as to their meaning. If only one is left, he can assign it to its definition in spite of absolute ignorance of its meaning.

Cause and effect. One of the most useful ways in which matching may be employed is in relating causes to effects. Barr in his history test has this item:

> In the left-hand (first) column below is a list of causes. In the right-hand (second) column is a list of results. Take your pencil and draw a line connecting each cause with the proper result.
>
> (a) The sinking of the Maine
> (b) The growth of industrial combinations
> (c) The annexation of Texas
> (d) The Dred Scott decision
> (e) The Lewis and Clark expedition
>
> (a) Territorial claims on Oregon
> (b) The Mexican War
> (c) The growth of abolition sentiment
> (d) Spanish-American War
> (e) Sherman Anti-Trust Law

You can easily get from this exercise an idea of the possibilities of the matching device in the matter of cause and effect. Barr's test was published in 1920. If he were publishing it now he would probably abandon the clumsy device of drawing lines between items. The same idea might have been more neatly carried out by leaving the parentheses empty in front of each item in one column and by requiring the pupil to write in these empty parentheses the letter or number of the corresponding item in the other column.

Other relationships. Whenever you wish to test a pupil's ability to put together pairs of things which belong together, you may do so by the matching device. In geography you may wish him to associate countries and products, cities and famous institutions, industries and physical conditions, cities (or other "features") and the countries to which they belong. In the case of countries and products, one way is to give the pupil a list of countries and a list of products.

One list should have a blank space before each item. The other should have each item numbered. The pupil is to put the numbers from one list into the blanks in the other. In history you may wish to test a pupil's ability to relate pairs of men who belonged to the same period, or pairs of men who were associated with the same event, or pairs of men who were opponents in some connection. You may likewise wish to ask pupils to relate men not to each other but to the events with which they were identified—Hamilton to the United States Bank, Clay to the Missouri Compromise, and Morse to the invention of telegraphic communication. Finally, you may wish to bring into your testing the relationship of events to each other—say, in point of time. One way of doing this is to arrange two lists of events and, by the device already suggested, to have the pupil relate each item of one list to its corresponding item in the other. For example:

Write in the front of each event in List I the number of the event in List II which occurred at about the same time.

List I	List II
(3) Founding of Pennsylvania	1. Emancipation Proclamation
(4) Election of McKinley	2. Battle of Lexington
(5) Lewis and Clark expedition	3. Opening of the Mississippi Valley by LaSalle
(2) First Continental Congress	4. Spanish-American War
(1) Battle of Gettysburg	5. Purchase of Louisiana

Combinations. One may combine some of the techniques which I have already suggested. For example, the completion method may be changed from a *recall* type of testing to a *recognition* type by combining it with the matching method. There is, as I have suggested, considerable difficulty in rating completion tests unless they are carefully made. The pupils are likely to write into the blanks words which the examiner has never thought of. Sometimes these unforeseen responses are entitled to at least part credit. Accordingly, the matching idea has been used. A list of words is furnished along with the sentences containing blanks. The pupil is to fill the blanks by using words taken from the list.

The following might be part of a test in plane geometry:

> In the following passage some words are left out. Find these words in the list to the right. In each blank in the passage write the number of the word, taken from the list, which belongs in the blank. The first two blanks are filled in to show you how to do it. Notice that the words in the list are arranged alphabetically and that there are more of them than you need.

The portion of a __8__ surface bounded by a __7__ called the circumference all points of which are ______ distant from a given ______ is called a circle.

A ______ line drawn through the ______ and terminating in the ______ is called a chord. An ______ is a portion of the circumference thus cut off. The part of a circle included between two ______ and an ______ is called a ______.

1 arc
2 center
3 circle
4 circumference
5 curved
6 equally
7 line
8 plane
9 point
10 radii
11 secant
12 sector
13 segment
14 straight
15 tangent

The purpose in combining the matching with the completion device is to convert the latter from a recall to a recognition type and thus to make it more definite. Sometimes, however, a combination of testing devices is made, because the examiner desires to *do two things* with each question. For example, Charters in his Language and Grammar Tests and Kirby in his Grammar Test wish to test the pupil's knowledge both of usage and rules. Usage lends itself to the alternative-response device and the application of rules to the matching device. Accordingly, Kirby presents in one column sentences in which an incorrect word is to be crossed out and in another column a set of rules from which one is to be selected and applied to each sentence. In each of the following sentences the pupil is to cross out one of the words in parentheses, and then in front of the sentence he is to write the letter of the rule that applies:

Sentences	Rules
( ) 1. It is (I) (me).	a. A predicate noun or pronoun is in the nominative case.
( ) 2. (Whom) (who) did you call?	b. A pronoun agrees with its antecedent in number.
( ) 3. We are the ones he called, you and (I) (me).	c. An appositive modifier is in the same case as the word it modifies.
( ) 4. A person should be careful of (his) (their) reputation.	d. The direct object is in the objective case.

Another interesting combination method shows the application of the new-type examinations to testing in a foreign language. In the Godsey Latin Composition Test somewhat the same idea is used as in the English test from which I have just quoted. Correct usage and rules are tested. There is, however, this difference: whereas in English there are only *two* ways in which an English-speaking pupil is likely to say a thing, there are in Latin *several* ways. Perhaps the Roman child made standard errors, so that if Quintilian had hit upon new-type examinations—new in 75 A. D.—he might have used the alternative-response method. Modern children, however, have so many wrong ways of expressing an idea in Latin that the Godsey test uses multiple choice. Moreover, it uses a second multiple choice to bring in the proper rule of grammar. I give below two exercises. The pupil is to encircle the right Latin form in the parentheses—the first multiple choice—and to encircle one of the four numbers. These numbers refer to a list of rules (not given here) printed at the bottom of the page. This, then, is the second multiple choice.

(a) *The king's sons are leaders of the enemy.*
Rēgis (fīlius, fīliōrum, fīliōs, fīliī) ducēs hostium sunt ....4..7..13..14

(b) *They were hurling weapons at the wall.*
Tēla in mūrum (coniciēbantur, coniciēbant, coniciunt, conicerent) ....1..8..13..14

Note: Rule 13, "the subject of a finite verb is in the nominative," applies in (a), and Rule 14, "action thought of as going on in the past is denoted by the imperfect," applies in (b).[13].

RECAPITULATION

In the foregoing pages I have considered the so-called new-type examinations according to the following scheme:

[13] Godsey, Edith R. *Latin Composition Test.*

- I. The recall examination
  - 1. Question-answer
  - 2. Completion
- II. The recognition examination
  - 1. Alternative response
    - a. True-false
    - b. Others
  - 2. Multiple choice
    - a. Information questions
    - b. Thought questions
      - One selection
      - Two or more selections
    - c. Definitions
    - d. Synonyms
    - e. Cross-out
    - f. Similarities
    - g. Analogies
    - h. Best reasons
    - i. Pictures
  - 3. Matching
    - a. Definitions
    - b. Cause and effect
    - c. Other relations
  - 4. Combinations

HOW TO MAKE NEW-TYPE TESTS

It is evident from the foregoing that the movement for better testing as it relates to new-type examinations has already covered a great deal of ground. The literature on the subject is extensive, and it is growing rapidly. I have given a few of the more important titles in a bibliography at the end of this chapter. The teacher who would be regarded as truly expert must have these techniques at his disposal. In conclusion I should like to give for the benefit of the teacher who may wish to try some of these techniques in his own work a few general directions as to procedure.

Do not suppose that yours is the one subject to which these objective tests do not apply. I have heard teachers of every subject taught in the public schools declare these examinations might be all right in other subjects but that in their particular subjects they could not be used. They can be used if you have the patience and ingenuity to work them out.

It should be said at the outset that you cannot make one of these examinations even after you have become somewhat experienced as quickly as you can write examinations of the "describe" and "discuss" type. It takes longer to make a multiple-choice test, or an alternative-response test, or a recognition-completion test than it does to construct the traditional type of examination. You will save time, however, in rating the pupils' papers and the more pupils you have the greater will be the saving. In other words, the time-consuming task is, in the case of the new-type examinations, shifted from the rating of papers to the preparing of the examination. *And that is where it belongs.* There is nothing productive, nothing creative in the rating of papers. Everybody recognizes it to be drudgery. With the old type of examination it is not only drudgery but drudgery largely wasted. The results are so unsatisfactory, so unreliable, that they fall far short of justifying the labor involved. On the other hand, the making of a good test paper is *worth doing*. It sets a standard both for one's self as a teacher and for subsequent classes—for these examinations so carefully constructed will, let us hope, be preserved by every teacher.

Before attempting to make a new-type examination, analyze carefully the teaching unit which you wish your test to cover. It may be the first case in percentage, or Brazil, or the Critical Period in American history, or magnetism, or the Elizabethan dramatists. Whatever it is, you should break it up into as many items as you may properly expect your pupils to be acquainted with. Then you should consider these items in their relation to each other and in relation to items in previous teaching units. You should also consider these parts from different points of view. Surely you will not be satisfied merely with information about them. You will want your pupils to invest them with richness of meaning, to show their sense of values with respect to them, to exhibit an appreciative attitude toward them, and to be able to think in reference to them. Incidentally, this searching analysis will be likely to show how you can improve your teaching of the same subject-matter on another occasion. This suggests that it would be a good plan to analyze the teaching unit *before* you teach it. Such an analysis would serve two purposes. It would improve your teaching, and with slight modifications it would form the basis of your testing.

After the analysis has been made, each topic and each idea about a topic may be made the subject of one or more short-answer key questions. I use the term "question" in a broad sense. I do not mean an interrogative sentence, but any test item requiring the pupil to do something.

As you proceed in the examination of your topics, you will doubtless find that some parts of the material lend themselves readily to the construction of true and false statements with which the pupils may be properly tested. Other parts may suggest a different kind of alternative-response test. There will probably be still other matters concerning which more than two choices may reasonably be set up. This will lead to a group of multiple-choice questions. Some of them may involve the definition of technical terms. Others may require the selection of ideas exhibiting a definite relationship to other items after the manner of the analogies test. Still others may ask for the selection of best reasons; and there will perhaps be a chance for a completion test in which key words will be omitted and set up in a column along with a number of other words. There may also be an opportunity for the matching of causes with effects—in geography the matching of imports with the countries from which the imports are received, or in history, of men and the events with which they are associated. No hard and fast rules can be laid down. You will have to experiment with them and no amount of detailed directions will save you that necessity.

Both experience and experiment have shown that special precautions must be taken to avoid ambiguities in wording the questions. For example, in making up a true-false test one must recognize that there are different kinds of truth and falsity. Some propositions are true by demonstration, others by preponderance of evidence, and still others by the weight of opinion. The falsity of statements is similarly conditioned. Moreover, a statement may be true or false according to certain conditions. In order, therefore, to make the statement universally true or false the conditions must be set forth in the statement. Even after care has been exercised in formulating the statements, it is desirable to submit them to competent people and to have these people actually work through the tests as the pupils are supposed to do. If this is done there will be few instances in which the writer of the examination will not find unsuspected imperfections in his work.

The new-type examination opens up rather extensive possibilities. McCall has estimated that more than twelve million examinations are given annually in the public schools of the country. I am not sure whether with the oncoming of the new techniques there will be more or fewer examinations, but I am sure that they will be better. The new-type examinations when compared with the old-type examinations will permit the average pupil to exhibit twice as much information about a topic.[14] Ruch estimates that for twelfth-grade pupils we may require, per minute of working time, any one of the following: five recall questions, six questions involving a choice among five items, seven questions involving a choice among three items, nine questions involving a choice between two items, or ten true-false questions. He says that this standard might well be cut in half for use in the elementary grades.[15] The new-type test is therefore rapid and searching. As Ballard says, it gets into the crevices and corners. It explores the entire area of the subject

Moreover, the result is very much more reliable. Ruch states that "ten- to twenty-minute examinations of objective type are very much more reliable than the five- and ten question traditional examinations which require thirty to sixty minutes."[16]

---

[14] Laird, Donald A. "A Comparison of the Essay and the Objectiv Type of Examination," *Journal of Educational Psychology*, 14:123- (February, 1923).

[15] Ruch, G. M. *The Improvement of the Written Examination.* Chicago Scott, Foresman and Company, 1924. 114 pp.

One of the inadequacies of the discussion test is the fact that it can hit only a few spots. Take the last test of this sort which you gave and after analyzing the teaching unit which it was supposed to cover, compare the topics touched upon in the test with those contained in your analysis. You will then realize that pupils with equal knowledge of subject-matter may, according to whether they are lucky or unlucky, obtain almost any sort of mark. If the questions in the old-style test happen to strike the "bare spots" in the pupil's knowledge of the subject, he will, of course, be unable to answer them. If they fall at the places where he is well informed, he will make a good showing. Sometimes in our rating of papers we actually raise the mark of a capable pupil who has turned in a poor paper. We assume he has been unlucky, although as a matter of fact he may have been quite unprepared. At any rate it would be very much better to have a test which told the truth.

Another inadequacy of the essay type of test has to do with the scoring. I have already pointed out that competent people may differ widely in rating such a test. I have also stated that a given person will rate the same paper differently under different circumstances. For example, toward the end of the rating of a set of examination papers I am by no means in the same frame of mind that I was at the beginning. The whole situation is different. Hence I do not rate the same. Again, a succession of good papers always makes the following paper which is not so good appear

[16] *Loc. cit.*

worse than it is. On the other hand, if the same paper followed a succession of bad papers, it might appear to be better than it is. The new method of examination meets the conditions I have suggested. Anybody, anywhere, at any time who knows the correct answer to each question can rate the papers. The pupils may rate them, and it is often a good plan to have them do so. A clerk may do so, and the possibilities of using clerks in schools to do the increasing, and important, clerical work of teachers is one of the things we must begin to think about. For all these reasons the new-type examination surpasses the essay examination in fairness to the student. The rating is the same for all. It is neither lenient nor severe; it is just. Moreover, since the new-type examination covers all the ground it should cover, indolent students may find it to their advantage to abandon their favorite pastime of guessing what the instructor will ask and devote their energies to studying the subject. The well-prepared student, on the other hand, need have no fear of misfortune.

The new-type examinations are generally conceded to be more interesting than the essay type of examination. In fact, nobody ever thought of the traditional examination as having any interest value. At this point we might well take a lesson from the experimentalists who have shown the superior worth of all work in which the worker is interested. Moreover, the new-type examinations are far less fatiguing than the examinations which they assume to supplant. In taking them the pupil has the sense of accomplishment, of progress. He is able rapidly to record the things he knows. In fact, in the older type of examination the tiresome part is the unessential part. The labor of writing for an hour or even two or three hours at breakneck speed in a race against time is both depressing and irrelevant. Success in examinations involving this requirement depends largely on handwriting and composition. Many a student, ignorant of the subject-matter with which the examination deals, "gets by" because of the facility with which he can write.

In view of these facts it is not surprising that pupils have invariably voted for the new-type examination whenever they have had a chance to do so. As McCall says of a particular kind of new-type examination: " 'Children cry for it' may be a bit exaggerated but at any rate they hate it less. It offers an opportunity for a contest where the rules are fair."

These new-type examinations—which I should prefer to call short-answer examinations—are being tried out in an experimental way by various examining bodies, for example, the United States Civil Service Commission, the College Entrance Examination Board, and the Board of Examiners for the New York City Public Schools. The classroom teacher should not be behind these organizations in willingness to experiment in this important field. Again, therefore, as so often before, we find opportunity for the teacher as a research worker. No findings of a civil service commission or of a board of examiners will be half as valuable as the findings of a considerable body of teachers.

BIBLIOGRAPHY

Asker, William. "The Reliability of Tests Requiring Alternative Responses," *Journal of Educational Research*, 9: 234-40 (March, 1924).

Ballard, P. B. *The New Examiner*, London: Hodder and Stoughton, 1924. 269 pp.

Barthelmess, H. M. "Reply to Criticism of Tests Requiring Alternative Responses," *Journal of Educational Research*, 6: 357-9 (November, 1922).

Batson, W. H. "Reliability of the True-False Form of Examination," *Educational Administration and Supervision*, 10: 95-103 (February, 1924).

Brinkley, Sterling G. *Values of New-Type Examinations in the High School, with Special Reference to History*. New York: Teachers College, Columbia University, 1924. 121 pp. (Teachers College Contributions to Education, 161)

Chapman, J. C. "Individual Injustice and Guessing in the True-False Examinations," *Journal of Applied Psychology*, 6:342-48 (December, 1922).

Chapman, J. C. *Trade Tests*. New York: Henry Holt and Company, 1921. 435 pp.

Colvin, S. S. "Marks and the Marking System as an Incentive to Study," *Education*, 32:560-72 (May, 1912).

Cooley, A. M. and Reeves, G. "Some Investigations Concerning the Use of Certain Home-Economics Information Tests," *Teachers College Record*, 24:374-92 (September, 1923).

Dickinson, Z. C. "Suggestions toward Improving Examination Marks," *Bulletin of the University of Minnesota*, 26: 31-36 (August 4, 1923).

Filer, H. A. and O'Rourke, L. J. *Annual Reports of the Chief Examiner and the Director of Research of the United States*

Civil Service Commission for the Fiscal Year Ended June 30, 1923. Washington, D. C.: Government Printing Office, 1923. 64 pp.

Fiske, T. S. *Annual Reports of the Secretary of the College Entrance Examination Board*, 431 West 117th Street, New York, 1921, 1922, 1923, and 1924.

Gates, A. I. "The True-False Test as a Measure of Achievement in College Courses," *Journal of Educational Psychology*, 12: 276-87 (May, 1921).

Hahn, H. H. "A Criticism of Tests Requiring Alternative Responses," *Journal of Educational Research*, 6: 236-40 (October, 1922).

James, B. B. "The Modern Test," *School and Society*, 19:209-13 (February 23, 1924).

Knight, F. B. "Data on the True-False Test as a Device for College Examination," *Journal of Educational Psychology*, 13: 75-80 (February, 1922).

Laird, Donald A. "A Comparison of the Essay and the Objective Type of Examinations," *Journal of Educational Psychology*, 14: 123-24 (February, 1923).

McCall, W. A. "A New Kind of School Examination," *Journal of Educational Research*, 1: 33-46 (January, 1920).

May, M. A. "Measuring Achievement in Elementary Psychology and in Other College Subjects," *School and Society*, 17:472-76 (April, 1923) and 17: 556-60 (May, 1923).

Miller, W. S. "An Objective Test in Educational Psychology," *Journal of Educational Psychology*, 16: 237-46 (April, 1925).

Monroe, W. S. *Written Examinations and Their Improvement*, University of Illinois Bulletin. Vol. 20, No. 7. Bureau of Educational Research Bulletin, No. 9 (October 16, 1922).

MONROE, W. S. "Written Examinations versus Standardized Tests," *School Review*, 32: 253-65 (April, 1924).

MONROE, W. S. and SOUDERS, L. B. *Present Status of Written Examinations and Suggestions for Their Improvement.* University of Illinois Bulletin. Vol. 21, No. 13. Bureau of Educational Research Bulletin, No. 17 (November 26, 1923).

ODELL, C. W. "Another Criticism of Tests Requiring Alternative Responses," *Journal of Educational Research*, 7: 326-30 (April, 1923).

PATERSON, D. G. "Improving the Examination Function in Teaching," *Bulletin of the University of Minnesota*, 26:47-56 (August 4, 1923).

PATERSON, D. G. *Preparation and Use of New-Type Examinations.* Yonkers, New York: World Book Company, 1925. 87 pp.

REMMERS, H. H., MARSCHAT, L. E., BROWN, A., and CHAPMAN, I. "Experimental Study of the Relative Difficulty of True-False, Multiple-Choice, and Incomplete-Sentence Types of Examination Questions," *Journal of Educational Psychology*, 14: 367-72 (September, 1923).

RUCH, G. M. *The Improvement of the Written Examination.* Chicago: Scott, Foresman and Company, 1924. 193 pp.

RUCH, G. M. and STODDARD, G. D. "Comparative Reliability of Five Types of Objective Examinations," *Journal of Educational Psychology*, 16: 89-103 (February, 1925).

SEASHORE, C. E. "College Placement Examinations," *School and Society*, 20: 575-78 (November 8, 1924).

TELFORD, FRED. "The Work of the Board of Examiners of the New York City Board of Education," *Public Personnel Studies, Bureau of Public Personnel Administration*, 2:268-87 (December, 1924).

Toops, H. A. *Trade Tests in Education.* New York: Teachers College, Columbia University, 1921. 118 pp. (Teachers College Contributions to Education, No. 115).

Weiss, A. P. "On Methods of Mental Measurement, Especially in School and College," *Journal of Educational Psychology,* 2: 555-63 (December, 1911).

West, Paul V. "A Critical Study of the Right Minus Wrong Method," *Journal of Educational Research,* 8: 1-9 (June, 1923).

Wood, Ben D. *Measurement in Higher Education.* Yonkers, New York: World Book Company, 1923. 337 pp.

Wood, Ben D. *The Reliability and Difficulty of the College Entrance Examination Board Examinations in Algebra and Geometry.* College Entrance Examination Board, 431 West 117th Street, New York, 1920.

Report of Commission on New-Type Examinations. College Entrance Examination Board, 431 West 117th Street, New York, November 3, 1923.

CHAPTER VII

GROUPING AND CLASSIFYING PUPILS

In the early days of the American school there were no grades. Schools were like the present one-room rural schools. In cities large rooms were provided in which several hundred children might be brought together under one teacher with a number of monitors or assistants.

THE GRADED SYSTEM

In the middle of the nineteenth century the idea of the graded school began to be expressed. This was particularly true in New England. Horace Mann spoke enthusiastically and frequently on the merits of the graded school. Some of his arguments are the very arguments which are now used for a much more accurate adjustment of the school to the pupil than the graded system ever permitted. It was believed that if one did no more than divide a school into eight or ten levels or grades, one could bring together pupils of the same ability. The idea was developed with conviction and with fervor. It was predicted that in a graded school the teacher of a grade would have to adopt but one method because the abilities of the children would all be the same. Reduction in effort, increase in efficiency, greater progress of pupils, and diminished cost of education were suggested as the probable results of adopting the graded system.

Well, the graded system was adopted—a milestone in the long road over which the American school has traveled. I have no disposition to minimize its importance. In its implications, if not in its actual work, it had in it all the things which Horace Mann saw. But straightway the graded system began to harden. It lost the flexibility which it might have carried over from the ungraded system; instead of facilitating the progress of pupils it often hindered it. It brought with it the promotion system and the "lock step" of the grades. The problem of the repeater—of the child who is left behind—likewise came into view.

Grades and promotion. As the graded system has developed, each grade has come to be thought of as a level of ability—ability being expressed largely in terms of the information which a child is able to give back when prodded by oral or written questions. The New England leaders seemed to think the pupils in a grade would be of about the same ability; and teachers, despite the evidence of their own experience, have continued to act as if they were. It would probably be wiser to regard a grade less as a stage in the advancement of pupils and more as a grouping of curriculum material. The gradation might then refer primarily to the offerings of the school and only secondarily to the ability of pupils. If the grades are thus regarded, the idea of promotion comes to be something quite different from the idea which is usually entertained and which indeed seems to be inevitable when the grades are thought of merely as stages of development. According to the usual idea, promotion is a certification that a child has successfully done the work of the previous grade. It would be an appeal to a higher and better law if promotion were based, not on what a child *has* learned but on what he *needs* to learn.

Consider, for example, the seventh grade as a grouping of curriculum material. This material is considerably richer than the material of the sixth grade. History may be taken up for the first time as a consecutive study. Shop work for the boys and domestic science for the girls may appear on the scene. There may be an opportunity to study a foreign language or general science. In English, "language" may give place to grammar. At any rate the seventh grade possesses greater opportunities than does the sixth grade. Who shall be permitted to take advantage of these opportunities? In general, of course, those who have done well in the sixth grade may properly be advanced to the seventh. But is this really the final answer to the question? Are there not others who should have this privilege? Is not the best interest of the child more fundamental than his attainment of seventy-five per cent of the work of the sixth grade? Is not his welfare more important in determining the issue than any per cent? May not a child who has been in the sixth grade for a year, even though he has done poorly, obtain more benefit from the richer curriculum of the seventh grade than he would by remaining another year in contact with the old curriculum?

Children in each grade are not of equal ability. As a matter of fact the chief assumption concerning the grades of the elementary schools is not true. The pupils in them are not of the same or of nearly the same ability. The evidence on this point is overwhelming. It may be drawn not only from educational literature but from every schoolhouse in the land. It is quite the usual thing for the pupils taught as a grade by one teacher to differ in mental age by four or five years. In extreme cases they will differ seven or eight years. If proper adjustments were made, each of these years of difference ought to correspond to a difference of one grade. Yet all these children are in the same grade. Moreover, if in a given school system all the pupils of each age from ten to sixteen are located, they will be found in every grade. You will find ten-year-olds, for example, in the first grade and you will find them in the eighth grade.

In achievement the range of individual differences within a given grade is likewise astonishingly wide. In spelling I myself found that a few third-grade children were better than average seventh-grade children and that on the other hand some seventh-grade pupils failed to do as well as average third-grade pupils. Others have found similar facts to be true in reference to other subjects. It has been said that in a given subject we can ordinarily take the best one-fifth of the children in a given grade, advance them to the next higher grade, and replace them with the poorest one-fifth of the pupils in the higher grade without changing the average ability of either grade. Under these circumstances it means little in reality, however much it may mean in theory, to say of a given child that he belongs to any particular grade.

Thus the graded system has failed to realize the purpose of its early advocates. It has broken down, in the sense that it has failed to bring together pupils of like ability.

TYPES OF CLASSIFICATION

I shall use the term "classification" in a broad sense to include three types of adjustment, namely, (a) segregation, (b) regrading (which may be considered under two aspects, regrading by subjects and absolute regrading), and (c) grouping within grades. Classification is, therefore, any arrangement of pupils which seeks to bring them together so that they may be instructed with the greatest benefit to each of them. It is an attempt to realize in the light of present knowledge and through present methods the benefits which in an earlier day were expected from the graded system.

Segregation. By segregation I mean the most extreme type of classification, namely, that which brings into one class children who are so greatly different from normal children that they cannot with benefit either to themselves or to normal children be instructed in regular classes. If a class for gifted children is organized such children are segregated. Usually, however, the term is used in connection with children who exhibit some form of defect. Thus we have many classes for the feeble-minded. In large cities we have classes for children who are defective in English and classes for children suffering from various types of physical defects. Under segregation should also be mentioned, in the interest of completeness, classes for adjustment, that is, classes the purpose of which is to bring up to grade children temporarily behind. Since segregation is an administrative measure, I shall devote no further attention to it in this chapter. For teachers of regular classes, segregation has no particular meaning except as a relief from the necessity of trying to teach one or more pupils who are entirely out of their proper element.

Regrading. When a pupil is changed from one grade to another in one or more subjects by a procedure other than that of regular promotion, we may call the procedure regrading. If the change from one grade to another is complete, that is, if it has reference to all subjects taught in the grade from which the change takes place, it may be called absolute regrading. Thus if, without qualification as to subjects, a pupil is advanced to the fifth grade from the fourth at a time or in a manner other than that of regular promotion, the act is one of absolute regrading. It would be the same if the pupil had been put back to the third grade. If a double promotion is conferred on promotion day, the extra promotion is an act of regrading. In fact, any change of grade outside of that contemplated by regular promotion is regrading.

Regrading may take place by subject. This happens so easily in the secondary school that it is hardly thought of as classification of any sort. Some elementary schools, however, recognize the varying aptitudes of pupils in different subjects. A fifth-grade child may be above grade in geography and below grade in arithmetic. These facts may be recognized by putting him in the sixth grade as to geography and in the fourth grade as to arithmetic. The act of making this kind of adjustment is one kind of regrading, namely, regrading by subject. The school which employs this method is sometimes said to be using subject classification.

Grouping within grades. The separation of pupils of a given grade into two or more divisions on the basis of ability may be called grouping within grades. Of the different kinds of classification employed by the school, this is the one which most closely concerns the teacher. It may be put into effect in either of two ways. If a school is large enough, a given grade may be divided into two or more classes each in a different room and under a different teacher. If these classes are made up so that they represent ability groups in the grade in question, then we have a very effective kind of grouping within grades, namely, *grouping by rooms*. In smaller schools, however, we cannot have more than one class of a given grade. In these schools if we think it is unwise to teach in one group pupils of such widely different abilities as the usual class presents, then we divide each class into two or more ability groups. This is another kind of grouping within grades, namely, *grouping within rooms*.

ABSOLUTE REGRADING

"Every day is promotion day for those who can benefit by it." Such is the slogan of many a school. Incidental promotion—that is, promotion within the year or semester —is always a possibility for the industrious pupil. This device is one of the prevalent ways of breaking the lock step of the grades. It may not be the best way but almost any way which really takes better account of individual differences among pupils is to be encouraged. "The lock step," said a superintendent recently, "is the goose step. I am done with it." Most of the adjustment secured by absolute regrading is in the nature of promotion. Relatively few adjustment cases of this sort involve demotion. The

Ohio Classification Committee[1] showed that according to the reports it received not more than fifteen or twenty per cent of regrading involved changing children to a lower grade. Accordingly, in what I shall have to say about regrading I shall frequently refer to it as a procedure for advancing pupils rapidly.

The opposition to regrading. The chief opponents of regrading are the teachers themselves. Many investigators have pointed out the reluctance of classroom teachers to permit children to be incidentally promoted or "skipped," as they say. McCall notes this in his book on *How to Measure in Education.* The common experience of administrative officers is to the same effect. Teachers do not like to lose their good pupils and it is, of course, the best pupils who are candidates for incidental promotion. Then there are those who argue against putting a child in the next higher grade—at other than regular promotion time—because, they say, he will be handicapped by the loss of certain work which he has skipped. There will be gaps in his information, they say. Some teachers honestly believe this to be the case. Others use the argument despite the fact that they know it to be fallacious.

The logic of the argument is attractive. On the assumption that a grade represents a level of development through which a child must pass as a prerequisite to the next level—on the assumption, in other words, of a rigid graded system—it *would* be detrimental to a child to skip a grade. But the grades are not what they purport to be. The curriculum of the fifth grade, for example, is not prerequisite to that of the sixth grade in any hard and fast way. Thousands of fifth-grade children are better than average sixth-grade children. It is no doubt true that the fractions of the fifth grade will be necessary for successful work in sixth-grade arithmetic. In geography, however, one does not need to study Europe before studying Asia. There is no sharply defined requirement in reading which a pupil must reach before he can profitably enter the sixth grade. The same is true of spelling and language. Moreover, one of the things we have lately learned is the surprisingly large amount of information children have about a given subject before they have received formal instruction in it. I shall have something more definite to say about this in Chapter IX. Arithmetic is practically the only subject in which there will be important gaps in a pupil's knowledge as the result of skipping. This is a small matter and can be taken care of by special arrangements both before and after the pupil is promoted.

---

[1] Buckingham, B. R., and others. *The Classification of Pupils in Elementary Schools.* Report of the Classification Committee appointed by Vernon M. Riegel, Director of Education. Columbus, Ohio, State Department of Education, 1925. 160 pp.

The fears on the part of teachers that pupils will not really profit by rapid advancement have so often been proven to be groundless that it is a wonder they have not been allayed. One investigator after another, one superintendent after another, bears testimony to the effect that when bright pupils have been given an extra promotion they have made good in the grade to which they were promoted.

Bases of regrading. A great deal depends upon the way in which regrading is put into effect. As things are now, even with so many standardized tests on the market, the prevailing basis on which regrading takes place is the teacher's judgment. In a recent investigation in Ohio, "teachers' estimates" was found to rank first in importance as a reason for regrading, while "class ratings" was next in importance. Class ratings are themselves a form of teachers' estimates. Only in about one-fourth of the reported cases of classification were standardized tests—including intelligence tests—the principal basis of action.

Yet test results have frequently been shown to be better than teachers' estimates. Wherever a thoroughgoing comparison has been made between tests and teachers' judgments, the former have proved to be the better. This has been particularly true in the matter of judging the intelligence of children, as I have shown at some length in Chapter III. The experimental evidence as it applies to the classification of pupils will be reviewed in a subsequent section of this chapter.

The most favorable times of the year. There is no time in the school year when regrading does *not* take place to some extent. January, however, appears to be the best month, with September as a second choice. It appears to make a difference whether the school in question is on the annual or the semiannual promotion basis. September is the favored month among schools that promote twice a year. There is evidence that the influence of a summer school is felt in this connection. Among the schools which promote once a year November is next to September as the month in which the greatest number of extra promotions are made. In fact, all through the fall months there is apparently a great deal of readjustment of the kind of which I am speaking, not only among the annual-promotion schools but also among those that promote twice a year. The late winter months and the spring months do not show much of this type of readjustment, unless the school in question is on the semiannual promotion basis. In that case March is a favored month. In short, there is evidence that in many schools there is a readjustment shortly after the general promotion has taken place. There is likewise a subordinate tendency in all schools to confer extra promotions in June. These are the double promotions—in other words, the advancement of pupils more than one grade at the regular promotion time.[2]

The most favorable grades. Regrading is considerably more frequent in the lower than in the upper grades of the elementary school. The Ohio Classification Committee's report shows that the greatest amount of regrading was in the fourth grade. There was nearly as much, however, in the first, second, and third grades, and the amount in the fifth grade was still considerable. Two-thirds of all the reported readjustments of this kind were in the first half of the eight-year elementary school.

Time saved. Among eighteen cities which reported to the Ohio Classification Committee the use of regrading there were 1,687 adjustments and these yielded a net gain in time amounting to 1,142 years. In other words, the saving in time due to regrading was the equivalent of a year of schooling for 1,142 pupils.

---

[2] All these statements are made on the basis of the investigation of the Ohio Classification Committee to which reference has already been made. The statements are made in positive form, and they are believed to be substantially correct. More extensive data, however, than the Committee had at its disposal are desirable.

Summing up as to the facts from the Ohio Classification Committee's report, we may say that the evidence points to the use of absolute regrading predominately in the lower half of the elementary school and to the utilization of this procedure as a secondary adjustment following the regular promotion period. The fall months are especially favorable for this action. Most of these adjustments involve promotion rather than demotion, and a large amount of time is saved for the pupils. The chief basis for decision as to reclassification is the teacher's judgment, either expressed as such or in the form of class standing. Independent evidence, however, is to the effect that tests, supplemented by the teacher's judgment, would constitute a more satisfactory basis.

The teacher's attitude toward regrading. No doubt the question of whether regrading shall constitute a part of the school procedure rests with the principal or the superintendent. In other words, *the policy* is not usually a concern of the classroom teacher. A large number of superintendents and principals express themselves in favor of the idea that every day shall be promotion day *for the child who is ready for advancement.* They say, however, that teachers are unwilling to put the idea into effect, that they are reluctant to say that the child *is* ready for advancement.

I think you should be open-minded on this matter. I admit that absolute regrading is a crude method of readjustment and that if a given school has for a long time been alert to avail itself of other types of readjustment, this one will probably seldom be used. But only a few schools have systematically availed themselves of any means of adjustment supplementary to those of regular promotion.

As I have already indicated, there is normally a wide range of pupil ability in any given class unless ability grouping has been put into effect. If, as I have already shown, there is a strong tendency for pupils to be able to do the work of the grade to which their mental age corresponds, then it must be true that normally there is considerable opportunity for a better adjustment through regrading. Moreover, the practically unanimous findings to the effect that capable pupils when given extra promotion make good in the grade to which they have been advanced shows that there is less hazard in regrading than may have been supposed. I cannot understand why teachers should be so conservative about this matter. They will readily admit that the few pupils to whom they have given extra promotion did good work in the higher grade. They seem to feel, however, that these pupils have missed something, although they are quite unable to locate it. They also have a vague notion about the nervous strain of the child and a general idea that it is undesirable to let pupils go through school rapidly. I should not recommend the extra promotion of a child who was not in good physical condition, but if this condition is present there is no evidence worthy of belief that a capable child suffers nervous or other ills because of extra promotion.

As to the general question of the wisdom or unwisdom of permitting children to go through school at more than the normal rate, the matter is so much one of opinion and experimental data are so difficult to gather, that I can give you nothing authoritative on it. Two or three facts, however, I should like to point out.

Physical versus mental maladjustment. In the first place most of the objection is due to the fact that these children will at an early age be graded with older children with whom they will be socially incompatible. This objection would largely fall to the ground if the practice of advancing children rapidly were more general. In other words, part of the remedy for the ills of skipping is more skipping. Moreover, the fact that the rapidly promoted child is unable in some respects to meet his older classmates on even terms does not settle the question. The real question is whether this admitted maladjustment is not, after all, an improvement upon the maladjustment which it supplants. For it is not a choice between satisfactory adjustment and maladjustment but between two maladjustments of different sorts. On the one hand, if we fail to advance the child he is maladjusted mentally in the grade in which we leave him. On the other hand, if we do advance him he is maladjusted physically in the grade to which we promote him.

Which maladjustment do you prefer? So far as I know, no one has anything to offer to help us in deciding this question. Could you not help in securing an answer? May this not be a research problem for you to consider? Could you not put through some carefully controlled extra promotions and observe the results, comparing them with the results secured before the promotion took place? At any rate the question is an open one. It is not one on which either you or I have sufficient grounds for a decision. Do not allow yourself to prejudge the matter, and when it comes to a question of advancing a really capable child, do not make him hurdle every barrier which you can raise.

Acceleration versus enrichment. In the second place, the question of acceleration versus enrichment of the curriculum is brought up in this connection. Regrading, as I am using the term, is by no means the only device in the public schools for advancing children rapidly through the grades—I am leaving out of account the relatively rare regrading which results in demotion. But the question of whether it is better to let bright children go through the grades rapidly or to maintain the same pace while providing a richer educative experience in each grade is a question which may as well be discussed here as in any other of the many connections to which it applies.

Let it be granted at the outset that a rich educational experience is what the schools should give. Theoretically it matters little whether that experience is given while the pupil spends a year in each grade or less or more than that time. It is entirely conceivable that a school may be so organized that it would be far better for the bright child to remain in the sixth grade than to go on to the seventh. The offerings of the sixth grade may be sufficiently diversified to permit him to get all and more than all that he could get in the seventh grade. But only a few schools are prepared to offer so diversified a curriculum with appropriately diversified teaching methods; and if a school *were* in such a position it would be organized that way with reference to the seventh grade as well as to the sixth. Even in such a school, therefore, it might be advisable to advance the child to the seventh grade, relying upon the rich curriculum of that grade to give him what he needed. The real point, however, is that very few schools are organized to do anything like the variety of work in each grade which the enrichment idea presupposes. Most schools furnish no more than a standard curriculum in each grade. Unless, therefore, a great deal of thought and planning, a great deal of high-grade leadership and high-grade followership have been present in a given school system, the enrichment idea falls to the ground. It is acceleration or nothing. Under these circumstances I plead for acceleration.

Another point is worth making. Acceleration *is* enrichment. If you skip a bright child from the third to the fourth grade and thus enable him ultimately to finish the elementary school, when without the extra promotion he could not have done so, have you not enriched his school experience? Suppose you and your colleagues in a given elementary school, whether by regrading or otherwise, bring a bright girl through the school in six years. Will not this tend to give her two years more of training in the high school or perhaps in college? It may mean that she will be able to take a normal-school course which her family might not otherwise have been able to give her. Under these circumstances, would not your action and that of your colleagues have been an enrichment of the girl's training? The educational experience in each successively higher grade is richer and fuller than that of any lower grade. If a policy of acceleration brings, as it inevitably will, this richer and fuller experience into the lives of more boys and girls, then the supposed antagonism between acceleration and enrichment vanishes. We have substituted for a present enrichment a later enrichment. We have chosen deferred dividends to present gain, and a tyro in finance will tell you which is greater.

REGRADING BY SUBJECTS

If a pupil is doing work in one or more subjects above or below the grade in which he is enrolled, he has received the benefit of regrading by subjects. Evidently this is a finer adjustment than absolute regrading. It recognizes the undoubted fact that children grow at different rates in different respects. At the same time, regrading by subjects is often a good introduction to absolute regrading. If a child has secured advancement to the next higher grade in respect to two or three subjects, it becomes easier to enroll him in the higher grade.

But in the case of regrading by subjects we have to consider to a considerably greater extent than in absolute regrading the demotion side of the question. According to the report of the Ohio Classification Committee nearly forty per cent of the pupils who were "displaced" on the basis of one or more subjects were displaced downward. Arithmetic was largely responsible for this. Sixty per cent of the children who recited arithmetic in grades other than their own did so in lower grades.

The subjects to which regrading is applied are predominantly arithmetic, reading, and spelling, although language, geography, and history are well represented. These statements are gathered from reports from eight city and four county systems. From the same source it is indicated that grades two to five inclusive are the ones in which such regrading is most extensively practiced.

From the very little that has been written on this subject, I judge that subject regrading is not a policy in many elementary schools. It is, of course, the rule in the high school and to a large extent in the junior high school. It is, however, entirely feasible in the grades. Suppose a fourth-grade pupil has satisfied his teacher as to all his fourth-grade work except language. If promotion is based on passing every subject, this pupil must be left back. Thus he may lose a year of time because of failure in but one subject. This happens even in supposedly progressive school systems. If regrading by subjects were recognized in the case of the fourth-grade child I have just mentioned, he would be promoted to the fifth grade and then required to do his language work in the fourth grade. The possibilities here in bringing him up in the subject in which he is deficient are at once evident. Again, suppose a fourth-grade pupil is particularly good in arithmetic but only average in the other subjects. It may even be admitted that as far as arithmetic is concerned he is "as good as anybody" in the fifth grade. He is evidently wasting his time reciting arithmetic in the fourth grade; but if subject regrading is not available we are obliged to let him continue to do so. In this case the child is caught in the machinery of the graded system.

How subject regrading may be applied. The matter is not only one of policy but of organization—neither of which is controlled by the teacher. It is entirely feasible to organize a graded school on the basis of advancement by subjects, provided the recitations in a given subject, say arithmetic all begin at the same hour of the day. When this provision is made, the sixth-grade child may go to the fifth-grade room at 9:45 for his arithmetic and do all the rest of his work with his own grade. There is a minor difficulty here due to the fact that the recitation periods are not the same length in all the grades. Suppose the period is forty minutes in the sixth grade and thirty minutes in the fifth. The program of the fifth grade could provide for some non-promotional subject such as music or physical training to follow the arithmetic. The sixth-grade child to whom I have referred could then either remain in the fifth-grade room during the extra ten minutes or return at once to his own room. In the latter case he could either listen to or participate in the sixth-grade arithmetic work. If there were several such pupils it would be worth while to organize the activities of this last ten minutes with reference to them.

GROUPING WITHIN GRADES

Largely a problem for teachers. As I have pointed out, regrading is essentially an administrative problem. The teacher is interested—vitally interested—but he does not ordinarily initiate the policy nor provide the organization. Moreover, even in working it out it is necessary for teachers to act coöperatively rather than individually, and this implies administrative and supervisory leadership. The grouping of pupils within grades, however, is more really a problem for the teacher. It is true that the question of policy will probably be decided as an administrative question, but the working of the policy is often in the hands of each individual teacher for his own classroom. I have already pointed out that grouping within grades may take place either by rooms or within rooms. When it takes place by rooms it is as much an administrative question as is regrading. Since grouping within rooms, however, is primarily a classroom question, I shall confine myself to that aspect of this type of adjustment.

Is grouping within rooms practicable? I lately had access to figures received from a score of cities in which this policy was employed. In these cities 117 schools did more or less grouping of pupils within rooms. This is a rather impressive, affirmative reply to the question, "Is grouping within rooms practicable?" Among these 117 schools this type of classroom organization was employed in every grade. It was used more frequently, however, in the lower grades. The practice was strongly in evidence throughout the first six grades, falling off sharply in the seventh and eighth grades due, no doubt, to the presence of the junior high school with its peculiar facilities for adjustment. Although grouping within rooms was common in the first six grades, it was somewhat more frequently employed in the first three or primary grades. In fact, the frequency was about one-third greater in the primary grades. This is not much more, if it is any more, than the larger enrollment of those grades. When, therefore, allowance is made for the larger number of pupils in the lower grades, it is not unreasonable to represent that grouping within grades is about equally applicable throughout the six-year elementary school.

Number of groups in each room. I have often been asked how many groups of pupils a teacher should make. As in the case of most educational questions, the answer is not singular but plural. The size of the class has something to do with the matter and the range of talent within the class has still more. The number of subjects in respect to which the grouping takes place likewise has bearing. No doubt in most cases the number of groups a teacher will organize will be the result of a compromise—a compromise between the larger number which the spread of talent would suggest and the smaller number imposed by the time schedule. If, for example, forty minutes of the school program is all you can allow for arithmetic, it is clear that you cannot make five or six ability groups, because the time you would have for each group would be too short to permit you to accomplish anything. You would lose too much time stopping and starting.

It will be of interest in this connection to know what teachers are actually doing. In the 117 schools to which I referred there were 784 instances of grouping. Of these, 359 were instances of two groups in a room and 357 were instances of three groups. In other words, 45.8 per cent of the cases involved only two groups and 45.5 per cent involved three groups. This makes up 91.3 per cent of the cases. Four groups of pupils in a room were reported 34 times out of the total of 784 cases. This amounts to a little over 4 per cent. All larger numbers of groups were scattering. If, therefore, the question of the number of groups to organize is to be settled by reference to the actual practice of teachers, it appears that the answer is "either two or three." So far as practice is concerned three groups are employed as frequently as two.

The number of pupils in each group. How many pupils you will put in each group will depend mainly upon the size of your class and the number of groups you make. The unity of the group is more important than its size. Teachers do not appear to make their groups of equal size. They tend to make the more capable groups larger than the less capable groups. The difference, although it amounts on the average to no more than two or three pupils, is nevertheless consistent in the reports I have received. This indicates a definite tendency to put fewer pupils into dull groups than into bright groups.

Whether this tendency is justified or not is perhaps a matter of opinion. One or two suggestions, however, may be made. In the first place, it seems likely that if children differ from the average to an equal degree they are equally in need of special and individual attention. This would doubtless mean that purely on the basis of adjusting the school to the child, the capable children should be given the same advantage as the less capable. But in the second place, the question of the instruction of bright and dull children is not wholly a matter of school adjustment. It is a social question. The value of the child to society both at the time of school attendance and later should likewise be regarded. From this point of view the school is justified in giving more rather than less careful instruction to bright children than to dull children. Many experts in education believe that the movement for the special instruction of dull children has robbed bright children of the attention to which they are entitled. On these grounds the advisability of teaching bright children in larger groups is open to question.

Bases of grouping. A great deal has been written on the subject of selecting pupils of different abilities for the purpose of grouping them. Most of these writings report the use of intelligence tests as the chief means of determining the ability of the pupils. I have been interested to find out, not from educational writings but from school superintendents and principals, the bases actually employed. I find that where the results of intelligence tests are available they are used, but that grouping as it is being conducted in the schools is carried out mainly on the basis of teachers' judgments. Among ten possible bases of grouping, the *teacher's estimate of intelligence* ranked first. The *teacher's marks* ranked second and the *teacher's estimate of industry* ranked third. The fourth item mentioned was *the pupil's age*. The fifth, sixth, and seventh ranks fell to *mental age*, *intelligence quotient*, and *standardized educational tests* respectively. *Previous school marks* and *health* were eighth and ninth respectively, and *results of individual intelligence testing* was tenth.

No doubt, as I have indicated, the rather low ranking of the intelligence and educational tests arises from the fact that in so many schools they are not given. For example, it is pretty well agreed that individual testing (for example by the Binet-Simon Scale) is highly reliable. The difficulty and expense, however, of administering individual tests prevent their being used widely. The fact, therefore, that the results of such testing occupy the last place in our ranking of bases of grouping no doubt merely means that this basis was seldom available rather than that it was thought to be of little value.

Since the teacher's judgment is so overwhelmingly important in connection with grouping (as well as in so many other connections), it is incumbent upon teachers to make every possible effort to refine their judgments. I have urged again and again that teachers should hold themselves to a higher and ever higher degree of expertness in their calling. An expert is a person whose judgment is keen, critical, and exact in his chosen field. Within that field his competence is admitted and his decisions are respected. I covet for the teaching profession the possession of the expert's capable judgment and the social recognition of the teacher as one whose opinion on matters falling within his field cannot be lightly disregarded.

GROUPING BY ROOMS

Wherever large numbers of pupils are available in a given grade, a superior type of grouping within grades is available. One may form sections or classes of the same grade and place each one in charge of a different teacher. If these sections are distinguished from each other on the basis of ability, there is a very evident opportunity to adjust the curriculum and the methods of teaching to the particular level of ability represented in each section.

The number of desirable ability groups is limited only by the number of classes of the same grade which may be brought together under one organization. Van Denburg[3] believes that when sixth-grade pupils enter the seventh grade of the junior high school the greater the number of ability groups the better. He gives reasons for this and his reasons seem to me to apply to other grades as closely as they do to the seventh. I have received reports of as many as twelve different ability groups in a given grade. Nevertheless, by far the commonest numbers of groups are two and three. In fact, eighty-nine per cent of the reports to which I have had access indicate no more than three groups per grade. Van Denburg[4] suggests a procedure to be followed in the seventh grade when grouping by rooms is available. His ideas are equally applicable to other grades. He points out that all the groups of a given grade will be substantially alike in effort and in thoroughness but unlike in achievement. There will be no repeating. There will be a chance for transfer between groups. Not oftener than twice a semester a uniform examination—that is, an examination the same for all the groups—will be administered. The "passing marks" will, however, be different from the different groups. In other words, we might be satisfied with 60 per cent for one group but would require 70 per cent for another and 80 per cent for a third. Van Denburg would set up in addition to these uniform examinations frequent review tests which would be of approximately the same difficulty for each group but quite different in the subject-matter covered. With respect to these tests, therefore, the same passing mark would be required in every group. If it were 80 per cent, each pupil would be expected to reach it in order to continue in his group.

---

[3] Van Denburg, J. K. *The Junior High School Idea.* New York, Henry Holt and Company, 1922. p. 45.

---

[4] *Op. cit.*, p. 45.

THE USE OF INTELLIGENCE TESTS

Although, as I have ~~shown~~, teacher's judgments of intelligence, proficiency, and industry are the most frequently used bases of grouping, the superiority of the results of intelligence tests for this purpose has been strikingly exhibited by several writers. For example, Alexander[5] reports that a first grade was divided into three sections during the first three weeks of school. The intelligence tests were then administered, and the pupils were reclassified strictly according to the test results. The author notes that some of the reclassifications according to the tests seemed at the time to be absurd but that later the tests proved to have been better than the teachers' judgments.

Doctors S. L. and L. C. Pressey[6] give evidence to the same effect. They report two investigations, one dealing with sixth- and seventh-grade pupils, the other with first-grade pupils. The pupils in the sixth and seventh grades had been sectioned according to teachers' judgments. On the basis of tests certain pupils in the slow sections made high scores and were accordingly transferred to fast sections. Correspondingly those in the fast sections who did poorest in the tests were transferred to slow sections. Eight out of ten pupils thus transferred to fast sections made good, and eight out of ten who were put in the slow sections were retained in those sections the next year. The conclusion is that teachers' estimates may be considerably improved if they are supplemented by test results.

---

[5] Alexander, Jessie M. "The Binet-Simon Test in Practical Use in the Public Schools of Hinsdale, Illinois," *Elementary School Journal,* 21:146-48 (October, 1920).

[6] Pressey, S. L. and L. C. "Measuring the Usefulness of Tests in Solving School Problems," *School and Society,* 12:531-34 (November 27, 1920).

The second investigation reported by the Presseys tells of the giving of an intelligence test to pupils entering the first grade and the subsequent *hypothetical* division of the classes into two groups. In other words, the children were not actually divided into groups except on paper. At the end of the first month of school the teachers divided their pupils into two sections on the basis of their judgments. According to the marks given the pupils at the end of the semester, the test grouping was superior to the grouping made according to the teachers' estimates in twelve out of the thirteen sections concerned in the investigation. In the remaining section the two groupings were equally effective.

Again Glass, reporting the results of grouping pupils in the Washington Junior High School of Rochester, New York, says that the method of division into ability groups up to 1919 was on the basis of elementary-school ratings. These, he says, proved unsatisfactory especially as regards the more capable pupils. He says in this connection, "Apparently the ability of teachers to detect low-intelligence pupils exceeds their ability to draw fine distinctions between very superior, superior, and high average groups."

Use both tests and teachers' judgments. Other writers give testimony to the same effect. The conclusion seems to me to be clear. It is that wherever possible the estimates of teachers should be supplemented by the results of testing. At the same time I would not have you suppose that the reliability of the test results is perfect. Several writers, Geyer[7] for example, have pointed out that in classifying pupils on the basis of intelligence tests we can expect that mistakes will be made. But the real question is, how much better is such a classification than the grouping which usually prevails? The point is made that a most noticeable improvement over the present practice may be obtained even though we fail to secure a high degree of accuracy. The position seems to be justified, as Stetson concludes—namely, that intelligence tests plus teachers' judgments yield better results than either of these alone. As I have taken occasion to say many times, testing will not and should not do away with the exercise of judgment. Why should it? After giving the tests, do we not still have available the teachers' judgments, just as we did before the tests were given? Unless these judgments are worthless we should use them.

ARE CLASSIFICATION AND GROUPING WORTH WHILE?

It is easy to see that bringing together pupils of similar ability ought to produce superior results. Moreover, there is some evidence to support this belief; but the evidence is by no means complete. We need more and better experimental investigation of this problem. In furnishing this experimentation the teacher may be especially helpful. Meanwhile let us look at some of the evidence which is now available.

---

[7] Geyer, Denton L. "The Reliability of Rankings by Group Intelligence Tests," *Journal of Educational Psychology,* 13:43-49 (January, 1922).

According to Moyer[8] superior and average pupils do somewhat better in ability groups than in mixed groups. Inferior pupils, according to the same investigator, do not appear to benefit by being placed together in one group. These conclusions are based on scores from standardized tests. No effort, however, appears to have been made to adapt the teaching and the curriculum to the pupils in each of these groups. Moyer's data likewise show that if pupils are grouped according to ability, those who are accidentally misplaced in higher groups than their ability warrants may actually profit by this misplacement, but that those who are misplaced in lower sections may suffer from such misplacement. It appears, then, that it is a more serious error to put a child into too low a section than into too high a section. The practical lesson to derive from this is: be sure you are right before you put a child into a low group. If you are in doubt, put him into an average or high group.

Torgerson[9] gives a rather convincing statement of the value of grouping and classifying pupils. At West Allis, Wisconsin, where Torgerson was director of educational measurements, classification had been in effect for three years when he made his study. He shows that failure was reduced during that period 41 per cent in the elementary grades and 19 per cent in grades seven to nine. This was not brought about through a reduction in the standards of promotion. These standards had in reality been raised during the period of classification. Evidence to this effect was secured from the use of standardized tests. These tests showed that grade achievement had been raised during the period of classification by the equivalent of more than one semester. The real difference, therefore, in favor of classification was even greater than the figures indicated.

---

[8] Moyer, Edward L. "A Study of the Effects of Classification by Intelligence Tests." *Twenty-third Year Book of the National Society for the Study of Education.* Bloomington, Illinois: Public School Publishing Company, 1924. pp. 313-22.

[9] Torgerson, T. L. "Is Classification by Mental Ages and Intelligence Quotients Worth While?" *Journal of Educational Research,* 13:171-81 (March, 1926).

Torgerson likewise makes the point that mental under-ageness and over-ageness were likewise reduced. The proportion of pupils who were three years above the mentality corresponding to their grades was reduced 73 per cent. Reductions of 55 per cent and 24 per cent were noted in respect to pupils who were mentally in advance of their grades by two years and one year, respectively. In making these reductions the procedure which in this chapter we have called absolute regrading was put into effect. Extra promotions were given to 346 pupils for a total educational gain of 409 semesters. "Not a single failure resulted, and in practically every instance the pupils thus accelerated not only did better work but were actually leaders in their classes." This is but a confirmation of the statement previously made, that capable children when given extra promotion are likely to make good.

It has been rather frequently found that bright children fail to work up to their capacity, while dull children do better work than their mentality would lead us to expect. Torgerson gives data from which he concludes that there is no evidence for the theory of an inherent lack of industry on the part of bright pupils. His data indicate that when pupils are properly classified and grouped the inverse relation so often noted between intelligence and accomplishment disappears. In other words, his results show that when pupils are given educational opportunity equal to their ability, bright pupils are as efficient as dull pupils. He even finds a tendency for them to have an advantage in this respect over pupils of the same mental age but of a lower intelligence quotient.

Glass,[10] to whom I have already referred, gives a good summary of the benefits to be expected from ability grouping. He is discussing the junior high school but his words have general application when he says: "Its chief virtue is that it demands a recognition of differences in individuals. It facilitates administrative plans in adaptation of curriculum requirements to individual capacity. It creates for each child an environment for the expression of his natural desire to compete with his peers. It avoids the discouragement of comparative incapacity and removes conditions tending to mental indolence. Since it serves to prevent, thereby, mental discontent, it becomes one of the main factors in the prevention of elimination. It stimulates the maximum of potential growth. It is an assurance of individual justice, and, therefore, an educational square deal."

Stetson[11] gives teachers' opinions on the merits of classification. Here are some of the expressions from teachers of bright sections: "Initiative has been encouraged and developed;" "students have been made more alert through the keen competition of capable classmates;" "responsibility for maintaining high standards of work has been felt;" "inclination to rest on past merit has been checked because maximum effort has been necessary to compete successfully with the group." Teachers of average sections gave these reasons: "They (the pupils) have been relieved of the fear of criticism from brighter pupils;" "they have developed more independence;" "thorough work has been possible at a comfortable speed;" "competition with equals has been enjoyed." Finally, the testimony of the teachers of poor sections includes the following: "They have escaped the ridicule of the capable students;" "their best efforts have been brought out and the best results of which they are capable have been procured;" "through small classes they have had the opportunity for frequent recitations."

---

[10] Glass, James M. "Classification of Pupils in Ability Groups," *School Review*, 28:495-508 (September, 1920).

[11] Stetson, Paul C. "Homogeneous Grouping in the First Year of a Five-Year High School," *School Review*, 29:351-65 (May, 1921).

ADAPTING INSTRUCTION

There is no guarantee in the classification and grouping of pupils that these advantages will be secured. Merely to segregate pupils or regrade them or group them within grades accomplishes nothing. These are devices which are intended to permit teachers to serve pupils better. A large part of the reason why evidence on the success of these devices has been so meager is that in too many instances the devices themselves, rather than the spirit with which they should be applied, have been in operation.

It is easy to divide pupils into groups in various ways, but it is by no means as easy to modify the curriculum and the methods of teaching in conformity with the ability of the groups. This is the point where the whole plan needs to be buttressed. It is also the point where teachers can play the most important part.

Bush brings out some of the points as to method in her report concerning a seventh-grade class of girls which was divided into two sections.[12] For example, the assignment in both sections for studying the geography of Norway and Sweden was: "Our history says the Northmen were brave and adventurous sailors and great wanderers. Can you find any geographical explanation of these qualities?" The A section gave very good reasons; the B section was helpless. Again, a history assignment was: "What did the Romans find in Britain?" The A section mentioned physical features, customs, and religion, as well as stories. The B section told only stories. In the use of reference books, in taking notes and in organizing material the A section showed its superiority. Its initiative was greater. Only the briefest suggestion or general question was required to start the A group in quest of knowledge. The B group required specific directions.

In dramatizing the Norman Conquest the A section caught the spirit of the event much more fully than the B section. A sand-table project, making an English manor, revealed some striking differences. The A section planned along general lines, neglecting many details. A number of the brighter pupils had the whole abstract plan in mind; but they gave no one a particular task which would help in carrying out the plan. On the other hand, the B section while lacking some of the general ideas, had thought more about details. Variation was likewise brought out in the planning and executing of a class party. The inefficiency of some of the children of high intelligence was evident. Their abstract intelligence seemed to help them little in a social way. The girl with the lowest intelligence showed remarkable ability in getting people to work together and to carry out her plans.

---

[12] Bush, Emma D. "An Experiment in Varying Method of Teaching Bright and Dull Sections of the Seventh Grade," *Journal of Educational Research,* 8:113-121 (September, 1923).

Stedman [13] gives some positive suggestions for the special treatment of children in the high groups. She tells of some of the things that were done with the gifted children in the training school of the University of California, Southern Branch. Such things as the writing of poetry; the writing of descriptive prose; detailed instruction in how to study, including the use of books and of libraries; instruction in foreign languages; opportunities to see and hear good music, plays, and lectures; trips into the country for the study of birds, animals, and growing things; opportunity to work out social and civics projects; definite training in leadership—these are only some of the things that can be done and should be done with children who are highly endowed. That such instruction pays society well is denied by no one.

The idea that I want to convey—and these articles will serve as a vehicle for doing so—is that differences in intelligence are both qualitative and quantitative and that the work must be adjusted with reference to each of these types of difference. Simply slowing up the work for the duller pupils and speeding up the work for the brighter pupils is insufficient. We must use different methods. Discovery as to these differences in method is going to come largely from the classroom teacher. Research therefore in the adaptation of method to different types and degrees of intelligence is particularly appropriate for teachers.

---

[13] Stedman, Lulu M. *Education of Gifted Children.* Yonkers, New York: World Book Company, 1924. 190 pp.

SUMMARY

The graded system was a great advance over the ungraded system, but its benefits have only been partly secured because as a system it has become rigid—an administrative device rather than a vital instructional force. Moreover, the graded system even as an administrative device is from its very nature insufficient. Accordingly, additional measures are needed to secure the best compromise between the economy of mass teaching and adjustment to the personal needs of pupils.

The purpose of classifying and grouping pupils is to bring together those who may be instructed in groups with the maximum of individual benefit. I have recognized three kinds of procedures representing progressive degrees of refinement. *Segregation* is for those who deviate very widely from normality. It cuts across grade lines. Segregated pupils differ so markedly from normal pupils that they may be handled together even though they differ in grade. The deficiencies, or other special conditions, of these pupils afford the unity which every instructional group must have. Reliance upon grouping within the grades and upon individual teaching will ordinarily suffice to take care of differences in grade. *Regrading* is generally belated adjustment. It is the resorting resulting from accumulated maladjustment and preliminary to a finer adjustment. In regrading, educational tests have especially evident uses.

Regrading is a commitment to "change of pace." Curriculum changes, enrichment, minimum essentials, etc., are not necessarily considered. In other words, regrading is feasible without reference to curriculum changes. In *grouping within grades* we reach the most accurate adjustment to the needs of pupils short of individual instruction. It is the last act of classification on the part of the administrator and in it the teacher's function likewise becomes prominent. The teacher is entitled to the benefits of segregation, regrading, and grouping within grades and should work to support these measures. For by them the teacher's task of reaching the individual is reduced to workable dimensions.

I have indicated some of the details which research has revealed in respect to these adjustment devices. A great deal more research is necessary, and in this the teacher should do his part. In particular we need to know very much more than we now know as to the best classroom technique for realizing the advantages which these procedures undoubtedly possess.

BIBLIOGRAPHY

ALEXANDER, JESSIE M. "The Binet-Simon Test in Practical Use in the Public Schools of Hinsdale, Illinois," *Elementary School Journal*, 21: 146-48 (October, 1920).

ARTHUR, GRACE. "An Application of Intelligence Tests to the Problem of School Retardation," *School and Society*, 10: 614-20 (November 22, 1919).

Evidence offered that "a child can, with rare exceptions, be made to work up to but not beyond the school grade corresponding to his mental age."

BUCKINGHAM, B. R. "Suggestions for Procedure Following a Testing Program: I. Reclassification," *Journal of Educational Research*, 2: 787-801 (December, 1920).

Bush, Emma D. "An Experiment in Varying Methods of Teaching Bright and Dull Sections of the Seventh Grade," *Journal of Educational Research*, 8: 113-21 (September, 1923).

Cole, L. W. "Prevention of the Lockstep in Schools," *School and Society*, 15: 211-17 (February 25, 1922).

Glass, James M. "Classification of Pupils in Ability Groups," *School Review*, 28: 495-508 (September, 1920).

McCall, Wm. A. *How to Measure in Education.* New York: The Macmillan Company, 1922. 416 pp.

Moyer, Edward L. "A Study of the Effects of Classification by Intelligence Tests," *The Twenty-third Year Book of the National Society for the Study of Education.* Bloomington, Illinois: Public School Publishing Company, 1924.

Pintner, R. and Nobel, Helen. "The Classification of School Children According to Mental Age," *Journal of Educational Research*, 2:713-29 (November, 1920).

Pressey, S. L. and L. C. "Measuring the Usefulness of Tests in Solving School Problems," *School and Society*, 12:531-34 (November 27, 1920).

Stedman, Lulu M. *Education of Gifted Children.* Yonkers, New York: World Book Company, 1924. 190 pp.

Stetson, Paul C. "Homogeneous Grouping in the First Year of a Five-Year High School," *School Review*, 29:351-65 (May, 1921).

Terman, L. M. "The Use of Intelligence Tests in the Grading of School Children," *Journal of Educational Research*, 1:29-32 (January, 1920).

Terman, L. M., and others. *Intelligence Tests and School Reorganization.* Yonkers, New York: World Book Company, 1922. 111 pp.

Torgerson, T. L. "Is Classification by Mental Ages and Intelligence Quotients Worth While?" *Journal of Educational Research*, 13:171-81 (March, 1926).

Van Denburg, J. K. *The Junior High School Idea.* New York: Henry Holt and Company, 1922. 423 pp.

CHAPTER VIII

ERROR AND FAILURE: THEIR EDUCATIONAL MEANING

A DISTINCTION is often made between *mass* teaching and *individual* teaching. In a very real sense there is no such distinction. Teaching corresponds to learning and learning is a personal activity. If there were mass teaching there would also be mass learning; and this is psychologically absurd. If it is individuals who learn, then it is individuals who are taught. There is no such thing as mass teaching. All teaching is and must be individual.

The measure of teaching, therefore, comes to be the summation of the learning, resulting from teaching, of the several individuals who are taught. It is the learning of pupil A, plus the learning of pupil B, plus the learning of as many other pupils as the teacher has under his instruction. This summation is made large or small by either one of two facts—first, by the learning of the several pupils and second, by the number of pupils. In other words, the larger the items the greater the sum, and the more items there are the greater the sum. In still other words, the more each pupil learns the greater the amount of teaching, and the more pupils there are in a class the greater the amount of teaching. Both these facts are important. If either of them vanishes, there is no teaching. If there are no learners, obviously there is no teaching. If all the pupils stay home, nothing is accomplished. But, on the other hand, a classroom may be full and no one may be learning anything. I have seen classes that appeared to approximate that condition. Here, too, the teaching is *nil.*

The measure of teaching, then, is the summation of the individual appeals which a teacher is able to make. It is the teaching of one, plus the teaching of another, plus the teaching of still another, and so on until all the persons taught by the teacher are included.

The problem is to make this individual teaching effective for as many pupils as possible. Economy requires that pupils be taught in classes or at least in groups. Accordingly, individual appeals must be made simultaneously. No teacher can rely upon teaching one pupil separately and then another and another and so on throughout the class. I remember that a university president gave as a reason for an upper limit of 25 in classes in Freshman English that there were 50 minutes in a recitation period. With 25 students he said the instructor could devote two minutes to each student, and he refused to allow the students to get less than two minutes apiece of the instructor's time. The only truth in this position was the fact that 25 is contained in 50 twice. The teacher who has a stimulus or a message of some sort for everybody in the class is the artist teacher; for the highest art in teaching is reaching the individual in the group.

The completeness with which this may be done evidently depends to a great extent on the character of the group with which the teacher deals. The size of the class is evidently important, although it is not nearly as important as most teachers suppose. It is surprising to what extent good teachers are able to do individual teaching—to have something for each pupil—in a large class. Such teachers are obviously of greater value to a school than teachers who must have small classes before they can produce good results. More important in its bearing upon the extent to which individual teaching may take place is the variety of talent present among the pupils in a group. A large class often prevents good teaching, not because of its size but because of the diversity of ability which it entails. Moreover, even a small class in which there are wide differences among pupils may present very great teaching difficulties.

Since the diversity of ability in a class has so great an influence upon the extent to which the teacher may reach the individual, a constant attempt must be made so to group and classify pupils as to permit an easy, frequent, and continuous individual appeal. In my judgment the teacher is entitled to the benefit of any administrative devices which will tend to unify the groups he teaches.

Individual teaching remains as truly the objective as it was in earlier days when the sons of aristocratic families were provided with personal tutors. Accordingly, I shall devote this chapter and the next one to the matter of individualizing instruction. In this connection the meaning of error and failure is especially important. Indeed, the whole question of school failure is fundamental. I shall, therefore, take up in this chapter the significance of error and failure.

THE TRAGEDY OF FAILURE

To err is human. Children make mistakes; teachers make mistakes; it is the common lot of humanity to err. All through the great literatures of the world runs the tragedy of failure. Shakespeare wrote four great tragedies. *Macbeth* is the tragedy of ambition, *Lear* of ingratitude, *Othello* of jealousy, and *Hamlet* of failure. It is not without significance that *Hamlet* is generally regarded as the greatest. What feeling is more poignant, what emotion is more pathetic or in its consequences more tragic than a great man's sense of failure?

Failure, then, in greater or less degree is the common lot. This idea is expressed in the theological doctrine of sin. We are all frail and imperfect, and the institutions we set up are no better than ourselves.

Failure in school means nonpromotion. The school is no exception. It has many shortcomings; but we do not speak of them as failures. We give the word "failure" a technical meaning. It refers to nonpromotion.

If we say that a pupil fails in a subject or in a grade, we mean that he was not promoted in respect to that subject or grade. We even take certain linguistic liberties with the verb "fail." One teacher says, "I failed him in algebra." Another says, "He failed algebra." The fact that "failure" has a special meaning in the school and that a colloquial usage has developed in respect to it is rather distressingly suggestive of a particular type of shortcoming.

Whose failure is this? Meanwhile, it is pertinent to ask, whether the nonpromotion meted out to the child is his failure or that of the school. The only basis on which we can say that the child has failed is the assumption that the school with its requirements is the standard which the child ought to fit and which he has failed to fit. This is not a tenable position. If we go deep enough into the matter we shall realize that there are no misfit children. This is the contention of Burk who did so much for individual instruction. There are misfit courses of study, misfit textbooks, misfit methods, and misfit teachers, but in the very nature of things there can be no misfit children. As well say a man does not fit his clothes as say the child does not fit the school. The child is the standard to which all other things must be adjusted; and to the extent that this adjustment is imperfect, to that extent the school fails. When, therefore, a child is not promoted, the failure, although it is usually attributed to him, is, in reality, the failure of something or someone else. It may be quite beyond the power of the school to prevent this failure. Sometimes the causes are so subtle and so difficult to cope with that the school as it is at present organized is obliged to accept defeat. While this may exonerate the officials of the school, it is, nevertheless, a confession of failure. In a broad sense, therefore, we may regard the failure of pupils in school as the failure, unavoidable though it may be, of the school itself.

PROMOTION AS AN ITEM IN EFFICIENCY

From this point of view the study of the promotion of pupils becomes a study, in part at least, of the efficiency of the school. The achievement of pupils—their knowledge and skill—may be a very misleading index of efficiency. A certain high-school principal, observing that the pupils who came from a neighboring elementary school were strong in English, sent the head of his English Department to study the methods employed in that school. He returned at the end of the first half day. He had learned the secret. His report was precise, accurate, and to the point. It was: "They kill them off." This elementary school should have had a graduating class of at least one hundred each semester. It had about thirty. Judged by the ability of these thirty, the school was a success, but it had failed twice as often as it had succeeded.

The effect of a given promotion rate. Ordinarily, 80 to 90 per cent of the pupils in a given grade are promoted. Let us consider the total effect of an 80 per cent promotion rate. Let us suppose that a thousand children entered the first grade of a certain school system last autumn and let us suppose in this school system annual promotions prevail. At the end of the first year 800 of these pupils will be advanced to the second grade. There they will join something like 20 per cent of the previous second grade to form a new second grade. Let us confine our attention to the 800 pupils promoted to the second grade and for the sake of simplicity, let us suppose that the 80 per cent promotion rate applies to these pupils as well as to the second-grade pupils who were left back.

Under these circumstances 80 per cent of the 800 pupils, or 640 pupils, will be promoted the following year to the third grade. The next year 80 per cent of the 640 pupils will be promoted to the fourth grade. The number is now only 512. This is, only a little more than half of the thousand who entered the first grade three years before. The destructive effect of the 80 per cent promotion rate is beginning to manifest itself. If we carry this group of children forward year by year, advancing 80 per cent of them at each annual promotion period, we shall find that only 168 of them will graduate from the eighth-grade elementary school in eight years. I do not mean that this is all of the original thousand pupils who will be graduated. Others will be graduated later and perhaps some will finish earlier. I merely mean that when the 80 per cent is applied successively each year, the result is 168 pupils who will be graduated at the end of the eight years.

If a 90 per cent rate prevails, 430 out of a thousand pupils entering the first grade will be graduated at the end of eight years. This may be compared with the 168 which resulted from the 80 per cent promotion rate. The promotion rate has been increased only one-eighth but the resulting number of graduates has been increased two and one-half times. If a 95 per cent promotion rate prevails (and such a rate is entirely feasible), 664 pupils will come through the elementary school in eight years or four times as many as was true when the promotion rate was 80 per cent.

We sometimes read of the astonishing sums of money to which a few dollars left at compound interest for a long time will amount. A similar effect is produced when the same promotion rate is applied again and again over a series of years. One fails to grasp the fact that this rate is applied to the same children year after year. Its effect is cumulative and it is astonishingly large. Merely to leave back one child in five on promotion day is not to end th matter. The school may have appeared to fail in but on out of five instances. This would be true if it went out o business at once. On the contrary, it continues year afte year to pyramid its failure. It piles up retarded pupils in the middle grades and eliminates them from the upper grades. This is failure on a large scale.

THE QUALITY OF REPEATED WORK

Why, then, do we allow pupils to fail? I have held many a teachers' meeting at which promotion lists were being discussed. With many variations in expression, the one reason given for failure to promote a pupil was that if he were to repeat the work of his grade "he would get it good." So far as I know there is no other legitimate reason. Pupils are required to repeat the work of a grade in order that they may do better work.

Some evidence on the question. Certain experimental data may be brought forward on this point. One investigation was made concerning one hundred pupils selected at random in an elementary school. Of these pupils, 26 had failed one or more times. The total number of failures was 31. In other words, there had been 31 instances of repeating the work of a grade. In only nine of these instances—less than one-third of them—did the pupils do better the second time they went over the work. McKinney, in connection with his investigation (unpublished) at Decatur, Illinois, and Springfield, Illinois, found that 35 per cent of the repeaters did better the second time they went over the work, that 53 per cent showed no improvement, and that 12 per cent did poorer work. Keyes[1] found that 28 per cent of the repeaters on whom he reported did better work, that 36 per cent did work of the same quality, and that 36 per cent did poorer work. No one of these items of evidence is very weighty, but you will observe that they agree rather closely. Their combined effect is therefore rather impressive. Apparently only about one-third of the pupils who repeat a grade do better the second time they go over the work. This tendency to improvement may be no greater than the added maturity of the pupils would produce. It may even be less than these same pupils would have exhibited if they had been promoted.

Recapitulation. In this question of promotion we have established a viewpoint which may be worth summarizing. In the first place, failure or nonpromotion is failure for the school rather than for the child. The effect of any promotion rate which is less than 100 per cent rapidly reduces the number of on-time pupils. It may easily make all efforts to reduce retardation and elimination of no avail. The usual reason for making a pupil repeat the work of a grade is that he may learn the work better. There is a consensus of evidence, however, to the effect that this expectation is not usually realized and the point is even raised whether the improvement which some pupils show may not be as properly attributed to an added year or half-year of maturity as to the fact of repetition. Finally, the fundamental question is raised as to whether any observed improvement is as great as it might have been had the pupil been promoted.

---

[1] Keyes, Charles. *Progress through the Grades of City Schools, a Study of Acceleration and Arrest.* New York: Teachers College, Columbia University, 1911. 79 pp. (Teachers College, Columbia University, Contributions to Education, No. 42.)

HOW TO INCREASE THE PROMOTION RATE

For these reasons teachers and administrators should consider ways of increasing the promotion rate. I am going to describe a rather simple plan as it was applied under experimental conditions by McKinney at Springfield and Decatur.[2]

At the end of the first semester of the school year 1918-19 the teachers in the elementary schools of these two cities made up their promotion lists as usual. On these lists 1,276 pupils were marked for failure. After the lists had been handed in, the teachers were directed to advance all their pupils but to put the pupils whom they had marked for failure on probation in the next higher grade. The probationary period lasted six weeks and during this time the receiving teacher kept close watch of the pupils who had been advanced on trial. At the end of the six weeks each teacher reported on these pupils, indicating whether or not they were maintaining themselves in the higher grade to the teacher's satisfaction.

Results. In the two cities 75 per cent of the 1,276 pupils who were put on probation made good according to the testimony of their teachers. The experiment was continued a second and a third term in these cities. In the second term 75 per cent of the probationers again made good. In the third term an epidemic of measles during the probationary period reduced the per cent to 63. In all three terms more than 2,200 children, nearly all of whom would ordinarily have failed, not only advanced to a higher grade but satisfied the teachers of the higher grades that they were entitled to stay there. The promotion rate in one of the cities—the record was not taken in the other—increased from 84 per cent to 95 per cent. At an 84 per cent rate a child's chance of being graduated on time from the eight-year school is about one in four. At a 95 per cent rate it is about two in three.

---

[2] Doctor McKinney has not published this report of this investigation, but I have summarized it in the May, 1921, issue of the *Journal of Educational Research.*

Recurrence of probation. It might be supposed that in a system of probationary promotion the same pupils would be on the probation lists term after term. This did not appear to be the case to any disquieting degree in the experiment I have just been describing. For example, of the 1,276 pupils who were put on probation the first semester, only 189 were put on probation either in the second or third semester of the experiment. In other words, 1,087, or 85 per cent of those who were probationers the first semester, were promoted without question the two following semesters.

Effect on scholarship. Another point which might be raised in connection with probationary promotion is the decline which it may be supposed to bring about in standards of scholarship. In my own judgment there is a danger in this question of standards—that is, rigid standards—which each child, no matter what his capacity may be, is required to meet. We should have standards to be sure but the best standards will be those which take account of the ability of the pupils. At Springfield and Decatur the general effect of probationary promotion upon scholarship as represented by the ratings of the schools may be indicated by the following statement: The average scholarship of the pupils of all classes in both cities for the semester just before the experiment began was 80.4. The corresponding figure at the end of the first semester of the experiment was 81.1. For the second semester of the experiment it was 82.2. Accordingly, the scholarship standards as shown by the figures themselves, instead of decreasing had actually increased during two semesters by 1.8 percentage points. This is not a large difference and it would perhaps be entirely sufficient to say that the standards of scholarship had not suffered. As a matter of fact, however, those who are acquainted with the nature of the averages will admit that a difference of 1.8 in a case of this sort, involving as it does fourteen or fifteen thousand children, is by no means a small difference.

A change in the point of view. Not the least important consideration in connection with probationary promotion is the shift of emphasis from the sending teacher and the old grade, on the one hand, to the receiving teacher and the new grade, on the other hand. According to the usual procedure, the sending teacher has all the responsibility for saying whether a pupil shall or shall not go forward. The receiving teacher oftens complains about the preparation of the pupils and this criticism is sometimes both outspoken and unprofessional. When, however, the sending teacher has nothing to say about promotion; when the promotion rate is based on the number of children who demonstrate that they can maintain themselves in the grade to which they have been advanced; when the pupil has a chance to show his ability on the job he has to do instead of being judged by his ability on a different job; and finally when the responsibility for making the decision as to promotion rests with the receiving teacher, we have provided not only for a new emphasis but also for a new result.

WIDER APPLICATION OF THE PROBATIONARY IDEA

Trial promotion is only one of the many ways by which the school may reduce failure. Segregation, regrading, subject promotion, grouping within grades, rapid advancement classes, double-track systems and the like are all designed to save the school and the child from failure and the consequences of failure. The probationary idea, however, is important in respect to all these devices. I have considered it in only one aspect, namely, its relation to the advancement of pupils at the regular promotion time, but the idea that a child should be given the chance to make good in any situation in which we contemplate placing him is well worth entertaining. Thus, probation may be applied in matters of segregation. We may give tests, consult records, interview parents, and supplement these activities by the use of our best judgment. But the real test of the efficacy of the placement of a child in a class for the gifted, or the feeble-minded, or the physically handicapped, is the response of the child to the new conditions. There may well be a period of trial before a final decision is made.

Similarly, the probationary idea may be applied to regrading. A pupil may be "tried" in the middle of the term, either in a higher or a lower grade. All the available information may and should have been secured, but nothing will settle the question so completely as the proof of the prediction in the situation to which it refers. You may say that the probationary idea is applied in your school. I have heard this claim made when all it meant was that if a promoted child did not make good he was put back. This is very different from having a definite and generally understood period—say six weeks—of probation, followed by a formal and recorded decision. There ought to be some point at which the trial comes to an end and the action which has been suspended is consummated or abandoned.

The probationary idea may likewise be applied to subject classification. We can try a child for four or six weeks in sixth-grade arithmetic or fourth-grade spelling. Again, we may give a child a preliminary trial in any group within a grade to which he may give evidence of belonging; or we may try him in a coaching class, if one is available.

All these arrangements for "change of pace" and for classifying children in various ways give flexibility to the graded school. If a probationary plan is added—and especially if this plan is applied so as to give the child every chance of progress of which he is capable—then the sum total of failure chargeable to the school may be materially diminished.

EDUCATIONAL MEANINGS OF ERROR

School marks. While it is true that failure is inevitable, it is also true that it affords an opportunity for improving conditions and a challenge to do so. Consider the matter of school marks. A pupil is said to have obtained 80 per cent in history, 90 per cent in spelling, or 75 per cent in arithmetic. I never yet heard a teacher tell a pupil or report to a parent the per cent of error corresponding to the per cent of success. It is clear, when one stops to think about it, that the "75" written on an examination paper or in a teacher's grade book or on the pupil's report card implies 25 per cent of failure or error. Now, the most effective way to increase the 75 per cent of success is through attention to the 25 per cent of failure. When, however, we fix attention upon the successes of the pupil without at the same time calling attention to his failures, we neglect the most direct and most powerful means of improvement.

The curriculum. Again, the educational significance of error may be illustrated in its application to the curriculum. A great deal of study is now being given to the offerings of the school. Some of the best of this study is concerned with the analysis of errors. What to teach is closely related to what is not known. For example, when Charters wanted to take up a course in language teaching, he based it upon the errors which children made in written and spoken English. Again, the curriculum in spelling—that is, the words one needs to teach—is decided in part by reference to the words which children have difficulty in spelling.

Teaching emphasis. There is, likewise, the question of emphasis in teaching. This involves not only the care and vigor with which the first teaching of a subject is done, but also the amount and distribution of drill which the subject may require. Some very competent investigations have been made as to the difficulty of the number facts—one measure of difficulty being *frequency of error*. Thus Heilman, Holloway, Smith, and Clapp have given us a very good idea of the difficulty of the addition facts. "Two and two are four" appears to be the easiest of all the facts. On the other hand, $7 + 9 = 16$, $5 + 8 = 13$, and in fact all the larger number facts are exceedingly difficult in comparison with the small number facts. Accordingly, the emphasis in teaching should be on the larger facts, not because they are larger but because the errors children make in them are numerous.

Many of the analyses of teaching materials would be greatly improved by attention to the matter of error. When the importance of items in history or geography is decided according to the number of times these items are mentioned in current magazines, there has often been a failure to distinguish between what is important to know and what is important to teach. Many of the commonest of the items do not need to be taught at all, at least not in school. Life outside of school will take care of them. Again, in the matter of job analysis, when a particular occupation or activity is subdivided into subordinate activities, many of these are of no consequence in a program of teaching because they present no difficulties. People do not make mistakes in doing them.

Failure, therefore, properly regarded, presents a chance to improve the work of the school. It is an opportunity and a challenge, not something merely to be accepted and bemoaned. Accepted it must be to some extent, for perfection is seldom attained. But the meaning and the philosophy of failure may be very stimulating.

TRIAL AND ERROR LEARNING

The relation of error to the work of the school may be looked at from another point of view. No doubt you have heard of "trial and error learning." This is the way animals learn. If you put a cat in a box or labyrinth in such a way that he must overcome some obstacle, such as finding the right path or opening a door, before he can obtain food, he will try first one way and then another until he gives up or hits upon the right way. If he succeeds, he will find the right way more quickly the next time we put him in the same situation. Ultimately, if the task is within his learning powers he will take the right way without trying any of the wrong ways. Some human learning takes place in this way, but in the training of young people for adult living trial and error learning is too wasteful. The school, therefore, seeks to shorten the method, in other words, to reduce the number of errors. We do not let the child learn that 7 and 8 are 15 by permitting him to come to grief through experience with other numbers than 15 as the supposedly correct answer. We put him in a situation where the correct answer is furnished and we take precautions to prevent a wrong answer from being mentioned or shown. Thus we at once associate the situation "7 and 8" with the correct answer "15." All erroneous responses are ruled out.

Thus the school short-circuits the trial and error method of learning. It forestalls error by methods known to be successful in so doing. It recognizes that the "natural" method is the wasteful method. It is clear, however, that if errors are to be ruled out we shall be helped by knowing what they are. Moreover, it will be desirable to know the ones which are most likely to occur. With respect to these probable errors it will also be advantageous to know the reason why they occur. Many times a knowledge of the cause of an error will help us to prevent its recurrence.

THE ANALYSIS OF ERROR

There are probably no new errors. All the varieties you have ever seen during your teaching career have almost certainly been made before. They have definite causes and these causes are not only common but rather few in number. Accordingly, the errors which pupils make may be classified into types.

In reading. Thorndike[3] did this in connection with a test in reading. Several paragraphs were given to about 500 children and each paragraph was accompanied by several questions. The test therefore required the child to put himself in the attitude of reading a paragraph and answering questions about it. One type of error arose from the fact that a pupil got the idea of writing something, but failed to get the reading idea or at least to connect these two ideas. Thus, as an answer to the question, "What did Tom give the girl?" a child wrote, "Said Mary." Again, some children were attracted by the question and apparently paid no attention to reading the paragraph. These children, unlike the child mentioned above, gave an answer which, under other circumstances, might have been a true response.

---

[3] Thorndike, Edward L. "A Study of Errors in Reading," *Elementary School Journal*, 18:98-114 (October, 1917).

For example, to the same question, "What did Tom give the girl?" a child wrote, "Tom gave the girl a skate." The answer according to the paragraph is, "A black dog."

Another type of error occurred when a pupil took material from the wrong paragraph. A question belonging to Paragraph A was, "Who gave the cat to Mary?" Some children answered, "Tom," although Tom is the boy mentioned in Paragraph B. Other children were attracted by the wrong item in a paragraph—an item which had no relation to the question. Children who made this type of error apparently used any word from the paragraph which made sense. A similar type of error arose when some part of the question itself attracted too much attention. For example, in the question, "What did the girl give Tom?" the word *girl* acted to produce the response, "Mary." Finally, just as there were types of error due to the "over-potency" of words or phrases in the paragraph or question, so there were types of error due to the "under-potency" of such words or phrases. For example, certain necessary qualifying words were left out of account. A child wrote "Yes" to the question: "Must a good boy always do all his school work?" The answer was wrong as a due account of certain phrases in the paragraph would have shown.

Thorndike notes in this connection a tendency which we have all noticed for pupils to accept as satisfactory whatever ideas arise quickly. Among pupils of all ages wrong answers are too frequent in comparison with admissions of ignorance. We ought to rate a pupil's work so that these admissions should not be penalized as heavily as an absolute or especially harmful error. This is a very suggestive point. I am reminded of the story told by Myers [4] of the observation of a class for a month to see how many pupils would say, "I don't know." During that time no child ever said it.

Thorndike notes still other types of errors in this "read and answer" situation. Children "fish" around for an answer and use it uncritically. The word "blue" is in the paragraph, so the dog's color is given as blue. We have all noticed this type of answer. If a child can quote something from the textbook, he is often quite content, although his quotation may have nothing to do with the question. Thorndike thinks that the remedy here is not to suppress the quotation but to subject it to criticism to see if it meets the need.

Other subjects. Several studies have been made of errors in spoken and written English. I have already spoken of Charters' work.[5] Similar studies—such as that of Cornman [6]—have been made in spelling. The investigation of errors in arithmetic has occupied the attention of several people—Uhl, Counts, Morton, and Osburn. Osburn's work [7] is particularly interesting. It relates to primary arithmetic. Since in making his analysis Osburn had access to more than 21,000 errors, his work is particularly authoritative. I shall refer in detail to some of his findings, partly for their intrinsic value and partly to illustrate the possibilities of the analysis of error.

---

[4] Myers, Garry C. *The Prevention and Correction of Errors in Arithmetic.* Chicago: The Plymouth Press, 1925. p. 26.

[5] Charters, W. W. "Minimal Essentials in Elementary Language and Grammar," *Sixteenth Yearbook, National Society for the Study of Education, Part I.* Bloomington, Illinois: Public School Publishing Company, 1917. pp. 85-110.

[6] Cornman, Oliver P. *Spelling in the Elementary School.* Boston: Ginn and Company, 1902. 98 pp.

[7] Osburn, W. J. *Corrective Arithmetic.* Boston: Houghton Mifflin Company, 1924. pp. 30-39.

AN ERROR ANALYSIS IN ARITHMETIC

Addition. A number of pupils gave 90 as the answer to 70 + 29 = ... The trouble was in adding 9 and 0. The pupils called it 0. The zero combinations give trouble in arithmetic. The type of error in which the operations are confused is rather common. Multiplication is sometimes confused with addition. For example, according to Osburn's analysis, 88 children, some of them in the sixth grade, gave "4" as the answer to 4 + 1 = ...

Sometimes pupils know what the operation is but do not know how to perform it. Suppose one of your pupils gave 15 as the answer to 47 + 22 = ... Can you see how he got it? or, did he just put down what came into his head? When you learn that this answer was given by about three per cent of Osburn's third-grade children, you begin to suspect that your pupil's answer is not as highly original as you may have supposed. He got it by adding the digits (4 + 7 + 2 + 2 = 15). Even if he had "just put down what came into his head" the error would have been a recognized type. It would have been due to an over-potency of the idea of writing an answer—much the same as was the case when the child, reported by Thorndike, wrote "said Mary" in answer to the question, "What did Tom give the girl?"

On five examples in column addition Osburn notes that 597 third-grade children gave one hundred wrong answers.

Each of these answers had a cause, and the discovery of these causes is the key to success in teaching these children to add columns correctly. For other third-grade children are daily making the same errors for the same causes.

Subtraction. A certain number of children, with the idea that in subtraction you must always take a small number from a larger one, will give 17 as the answer to 11 — 8, or 14 as the answer to 15 — 9, or 243 as the answer to 675 — 438. Forgetting to borrow is common, as is shown when a pupil gets 180 for 470 — 390. Trouble with the zero combinations is the cause when a pupil gets 30 for 60 — 35.

Multiplication. The same trouble in handling zero is shown by the pupils who give 314 or 1244 as the answer to 310 × 4. You can depend upon these answers turning up among a few pupils—say four per cent of those in the third grade and perhaps two per cent of those in the fourth —if you give this example. This statement is, of course, made on the supposition that no more than the ordinary care has been taken in the treatment of zero. When a certain number of children give 9408 as the answer to 1076 × 8, do you recognize this error and can you give its cause? How about 9455 as the answer to 155 × 60?

Division. In division combinations, typical errors were: 1 as the answer to 1 ÷ 6; 18 as the answer for 6 ÷ 3; 5 as the answer to 0 ÷ 5; 0 as the answer to 4 ÷ 4; and 16 as the answer to the same combination. In the example, 56042 ÷ 14, 57 children in grades five and six got either 404 or 43. The trouble in these instances was either using the wrong operation or difficulty with zero.

Fractions. In fractions no end of children will say that $1/2 + 3/5 = 3/7$ (adding the numerators together and the denominators together). Others will say that $2/3 + 3/5 = 6/15$ (multiplying the numerators as well as the denominators). Still others will give $1/18$ as the answer to $1/6 \times 3$.

All teachers of arithmetic will recognize these as typical errors. And, of course, there are a great many others. Some of them are very common, others occur only occasionally; but a study of the whole question leads Osburn to say, "there are very definite and very persistent type errors." And again, "it is idle to speak of these errors as chance errors." I would not have you believe, however, that the identification of types of error together with a determination of their causes is always as easy as it is in the cases I have presented. Some errors can be studied only when the teacher sits by the pupil and has him tell what he is doing as he goes along.

Long division. Osburn in a study of errors in long division made by five hundred children, finds the following facts to be true: (1) that 39 per cent of the errors arose from incorrectly estimating the quotient; (2) that 21 per cent arose from error in the multiplication combinations; (3) that 21 per cent were caused by error in the subtraction combinations; and (4) that 19 per cent of the errors were in borrowing, copying, carrying, bringing down, and failure to complete. More than half the trouble lay in multiplication and subtraction. In other words, these children were not ready for long division. Teachers are trying to teach this difficult subject to pupils who have not yet mastered the prerequisites.

Problems. Various analyses have been made of errors in verbal problems. A few general types cover a large number of cases. (1) Children use the wrong operation. "If we learn to spell two words a day, how many words do we learn in eight days?" The commonest wrong answers are 10, 6, and 4, indicating respectively addition, subtraction, and division. (2) Pupils likewise totally fail to comprehend. This is sometimes manifested by the absurdity of their answers. The child who gave the speed of a railroad train as 972 miles an hour failed to understand the conditions. So also did the child who represented that in a city of a million inhabitants, a certain health measure would save 3,333,333 lives. (3) Sometimes a pupil decides upon the right operation but carries it only part way. He loses confidence or encounters some obstacle. (4) Errors in the fundamental processes—that is, mistakes in adding, subtracting, multiplying, or dividing—are very common, as every teacher knows. Osburn found that about one-fifth of the errors made in solving his verbal problems were attributable to this cause.

Summary. There are no new or chance errors. On the contrary, errors run in types which can be identified by using a reasonable amount of ingenuity. These errors are found everywhere throughout the country. They show that the school is attempting the impossible unless it learns what these errors are. A very great improvement in the work of the school will result from the persistent investigation *on the part of teachers* of the best way to cope with these types of errors—to prevent their occurrence with the present distressing frequency, and to remedy them when they do occur.

THE SERIOUSNESS OF ERROR

Myers [8] has raised in a very significant way the question of the seriousness of error. In particular he seeks to find out what the influence of a mistake is upon the learner. He and his students recorded the errors made by pupils on number combinations. Once an error had been made by a child on a given combination, successive responses to that combination were recorded over a period of time. A certain pupil, for example, made the following series of responses to 4 + 6: 8, 8, 10, 8, 10, 10, 8, 10, 10, 10, 10, 8, 10, 10, 10, 10, 10, 10, 10, 8, 10, 10, 10. Observe that the initial error of eight, although it becomes relatively less frequent as time passes, can never with any confidence be said to have disappeared. Another child, responding on different occasions to 2 + 4 gave these answers: 7, 7, 6, 6, 7, 6, 6, 6, 6, 6, 6, 7, 7, 6, 7, 6, 6, 6, 6, 6, 7, 6. Again the wrong answer persists. The child is never sure that two and four are six. He has made a wrong association of two and four with seven and he never gets away from it. Here is a somewhat different series of responses. A third-grade child responding to 7 + 7 gives these answers in the order indicated: 41 (four times), 14 (seven times), 41 (twice), 14 (seven times), 41 (once), 14 (twice). The trouble probably is that the child doesn't know how to write fourteen. Possibly he does not at any time regard seven and seven as making forty-one. His error in writing persists throughout the series. At the end of the observation he has written the correct answer only twice since the last appearance of the wrong form. We can't be sure that the wrong form will not appear on the very next occasion.

---

[8] Myers, Garry C. *The Prevention and Correction of Errors in Arithmetic.* Chicago: The Plymouth Press, 1925. 75 pp.

From data such as these Myers concludes: "a mistake is immeasurably serious. Once it is made we are never sure it will not be made again in just that same way." Our nervous systems are full of the traces of old mistakes which rise up to confound us at the most unexpected moments. Are you sure of that column you lately added? Can you swear to the right answer without doing it over again? When you meet Jones do you have a persistent tendency to call him Smith? Did you ever get a wrong "set" as to the date when the Pilgrims landed, confusing it with the landing at Jamestown and never being *quite* sure whether it was 1607 or 1620? Myers is right. An error once made is never entirely lost. The nervous mechanism makes no distinctions. "A wrong answer in number work is apparently just as definite as a right answer. In terms of habit processes there is no difference between a wrong answer and a right one. Errors are not negative entities. They are just as positive as correct responses." I feel that I must enter a demurrer to this as a rather unguarded statement of what in its essentials is substantially true. There *is* a difference between a wrong answer and a right one in terms of habit processes if the satisfactions arising from the two responses are different. If the wrong answer is immediately associated with some unpleasant result, it is not as likely to recur as would be the case if satisfaction or a neutral feeling were present. I feel like inserting this caution because there is, in reality, a place in pedagogy for error. The old principle of learning by one's mistakes still holds *if the mistakes are at once evident as such and are associated with annoying consequences.* You can't more effectually teach a child to avoid the fire than to let him burn himself.

Substantially, however, and certainly in the sense in which Myers meant to be understood, all that he shows as to the seriousness, the recurrence, and the positive character of mistakes is sound doctrine.

THE PREVENTION OF ERROR

If error is so serious—if it always leaves its traces and its tendencies to recurrence—then the prevention of error becomes enormously more important as a school problem than the removal of error. Again let us consider primary arithmetic as the field from which illustrative ideas may be drawn.

Counting. I think primary teachers will agree that one of the most fruitful sources of error in arithmetic is the tendency of children to count. If a pupil has to add nine and six, it is found that the most frequent errors are the ones closest to fifteen. In other words, fourteen and sixteen occur with considerably greater frequency than any other wrong answer. The principal reason is that the child counts and either stops one short of the right answer or goes one beyond it. There are other reasons for not wanting a child to count besides the fact that he is likely to get the wrong answer. One reason is that the process is a slow one. Another is, that once it is established it is very hard indeed to substitute a direct response for it. Try, for example, to associate with the first word of each line of "America" the last word of the same line. When you think of "My" at the beginning of the first line try to associate with it the last word of the same line. See if you can do it without saying the whole line to yourself. The child who has learned to get the answer to 9 + 6 by saying "ten, eleven, twelve, thirteen, fourteen, *fifteen*" is in much the same situation that you are when you want to associate "thee" with "my," "liberty" with "sweet," and "sing" with "of." He has to run through the series. He may do it very rapidly but once having become a victim of this habit he cannot shake it off.

Do not suppose that because the child does not count on his fingers he is not counting. He may count with his toes, or by tapping, or by any other bodily action. Myers tells the story of a child who suddenly began to fail in his number work. It was found that he had lost a loose tooth which had hitherto played in effective combination with his tongue in counting.

If counting is so detrimental to success in number work, ways of preventing the habit should be carefully studied and applied. Some of our present methods directly encourage counting. When we teach 3 and 4 are 7 with objects, we not only fail to guard against counting from four up to seven but we actually suggest it. The trouble may easily be avoided. The teacher may first present the entire seven objects which the pupils must identify as a group of seven things. (At this point they may count if they need to; counting is to find out how many there are in a group.) The teacher may then divide the group into the required two parts, identifying (by counting if necessary) each part. The teacher may then recall the total number in the group and arrive at the statement desired. The vicious counting, "four, five, six, seven" has not been called for.

Again, our teaching favors counting because of the way we build up the series of combinations. We begin with one and one, two and one, three and one, etc. The idea is thus set up that the answer is just one more. Later when we deal with the two's, it is two more. Still later is becomes three more, and so on. A random order of presenting the number facts in addition would be better than this, although a random order is probably not the best order.

Primary teachers often have children resort to objects when they forget one of the combinations. The child is encouraged to get out his shoe pegs, or to go to a table where splints, cubes, and spheres are in abundance. This is considered in many quarters to be an excellent practice. It is in reality a bad practice. If, for example, a pupil resorts to objects to find out how many 9 and 6 are, he is almost certain to count from 9 up to 15.

I should prefer to prevent counting and guessing and the occurrence of wrong responses for any reason by posting on the walls of the room the full statement of each combination which the class is supposed to know. Moreover, I should encourage the children to refer to these combinations as often and as long as they wished to do so. The reprehensible thing would be to give a wrong answer when the right answer was in sight. If you adopt this device, you need have no fear that the children will continue to use the posted combinations longer than they should. They will drop them when they really know them, for the sufficient reason that it is more satisfying to give the answer than it is to take the trouble to find it where it is posted.

There are many other ways in which primary teachers can prevent the first occurrence of errors in number work. For example, they may set up in the class the ideal of saying "I don't know" instead of guessing. They may avoid interference in learning by keeping apart, in their teaching, combinations which have common elements, such as 4 + 5 and 4 + 3. The list of resources for preventing mistakes is probably a long one. The teacher can investigate this question with profit.

THE REMOVAL OF ERROR

Myers has pointed out that certain practices of the school encourage the recurrence of error. In our rating of pupils, omissions are counted the same as errors. I have already referred to Thorndike's idea that admissions of doubt should not be penalized as heavily as absolute error. Jim answers ten questions in a test. He knows seven of them and gets them right. The other three he does not know, but he guesses at them, getting them wrong. Alice likewise knows seven of the questions, but she omits the three she does not know. We give Alice and Jim the same mark. Indeed, Jim may succeed in getting a higher mark than Alice, because the chance he takes on the three questions he does not know may turn out in his favor.

Again, we are encouraging error by the nature of our emphasis on speed. Some of the standardized tests encourage this mistaken emphasis. If the number of questions attempted is scored independently of the number correct, there is danger that teachers and pupils may look upon speed as desirable apart from accuracy. No one is interested in the speed with which a child can get a thing wrong. Mere speed cannot be offered as a substitute for accuracy. A different manifestation of the speed mania occurs when courses of study are over ambitious in introducing new material. There are some lines of work in which it is a good plan to survey material rapidly and to go over it afterwards a second or third time, filling in details and making applications. In other subjects a slow approach is the correct approach. The practice of giving ten spelling words a day has all but disappeared in progressive schools. Two or three words a day—that is, new words which the child does not know how to spell—are quite sufficient. The same may be said as to the acquiring of a vocabulary in a foreign language. It is easy for the teacher to be deceived as to the number of words a child can learn. It is true that he can learn fifteen or twenty on any given day but he cannot keep it up without losing many of those which he had previously learned.

The question of how rapidly material should be taught is one about which we know very little. It is certain that present practice encourages errors among the less capable pupils because of its too great speed. On the other hand, it is likely that greater speed is desirable with the best third or quarter of elementary-school pupils. Since experimentation of this sort must be with the children themselves and must be carried out over a long period of time, the teacher is the only one who can undertake it.

Properly distributed drill will be effective both in preventing and in removing errors. I indicated in Chapter I that an item once learned should be repeated before forgetting sets in. I showed that this would mean repetitions which could be spaced at longer and longer intervals. It is too much to expect that the school can attain to anything like mathematical exactness in the distribution of repetition or drill. Gross mistakes, however, in this respect may be avoided. All of a pupil's relatively new acquisitions should be kept alive by occasional repetition. It is wise to do this even if new work cannot be taken up as rapidly as one would like.

Perhaps one of the most widely applicable principles for the removal of error is the law of effect. The error should be associated with dissatisfaction or annoyance. Most of the errors a child commits are not associated with anything. He is not even conscious of them. Think of the bushels of "home work" the children of a class turn in during the year. It is full of mistakes to which neither the child nor anyone else pays attention. Certainly neither home work nor seat work should be assigned unless the child is in full mastery of the required associations. This would rule out much written home work, especially work in arithmetic. It would, however, have little reference to home work involving the reading of books.

SUMMARY

For centuries society has been raising the requirements of adult living, until it is now unable to pay the cost of training for adult living on any but a mass-production basis. Yet in a sense there is no mass teaching. The class as a learner does not exist. Individual teaching is the only teaching.

Accordingly the reaching of the individual is the great art in teaching. One of the bases for doing this is a study of failure with a view to its prevention. I have therefore devoted this chapter to a consideration of failure and error.

Failure has a technical meaning in the school. It means nonpromotion. The promotion system with its eighty or ninety per cent promotion rate is the outgrowth of the graded system in virtue of which each grade is thought of as a necessary and sufficient prerequisite to the succeeding grade.

The destructive effect of a promotion rate when applied over a series of eight years is similar to the cumulative effect of compound interest. In the aggregate the number of pupils who will fail of graduation on time is surprisingly large. Moreover, a small increase in the promotion rate will produce a large increase in graduates at the end of eight years.

The merits of a plan of probationary promotion have been presented and the application of the probationary idea to all kinds of grouping and classifying of pupils has been suggested. By this means an even greater flexibility—an even more effective breaking of the lock step of the grades—is secured than is possible where grading, promotion, segregation, regrading, and grouping within grades are carried out without trying the pupil in the new situation.

It is recognized that failure is inseparable from human frailty. It is asserted, however, that error and failure afford endless opportunities for improving conditions. It is held that civilized society through its agent, the school, should derive every possible lesson from the existence of error. In school marks the school should scrutinize the margin of error rather than the "per cent correct" in order that the latter may be raised. A study of errors is fruitful in making up the curriculum, for the things to teach are the things in which people make mistakes. The application of error to methods of teaching is also pointed out. The things which are hard—that is, in which mistakes are frequently made—must be presented carefully and drilled frequently.

With more particular reference to the improving of instruction, the analysis of a pupil's errors is suggested. All the errors which a teacher is likely to encounter have undoubtedly been made before. Most of them may be classified and when this is done the remedy frequently becomes apparent. As illustrations of analyses of error, reference is made in some detail to investigations in reading and arithmetic.

The seriousness of error—its persistence and tendency to recur—is asserted; and the conclusion is reached that the prevention and removal of error is a duty which the school should assume to a greater extent than has hitherto been the case. Certain practices which foster error have been referred to. In all this the teacher's part is emphasized. After realizing the nature and seriousness of error—in other words, after having adopted what might be called a philosophy of error—one's first effort should be to prevent the occurrence of mistakes. A sincere effort should likewise be made to remedy the mistakes which one has not been able to prevent. The analysis of error and the study of types and causes of error will be the best means of making these efforts effective. They will enable the teacher to adopt accurately adjusted methods of procedure. Moreover, such work will provide a real opportunity for research on the part of the teacher. It will insure the attitude of studentship. It will enable the teacher to grow in service and to become truly expert.

BIBLIOGRAPHY

Barthelmess, Harriet M. "Geography Testing in Boston," *Journal of Educational Research,* 2:701-712 (November, 1920).

Buckingham, B. R. "An Experiment in Promotion," *Journal of Educational Research,* 3:326-35 (May, 1921).

Kallom, Arthur W. "Analysis of and Testing in Common Fractions," *Journal of Educational Research,* 1:177-92 (March, 1920).

McGraw, H. W. "Use of Test Data as a Basis for Drill in Grammar," *Journal of Educational Research,* 10:291-96 (November, 1924).

Morton, R. L. "An Analysis of Pupils' Errors in Fractions," *Journal of Educational Research,* 9:117-25 (Feb., 1924).

Myers, Garry C. "Persistence of Errors in Arithmetic," *Journal of Educational Research,* 10:19-28 (June, 1924).

Myers, Garry C. *The Prevention and Correction of Errors in Arithmetic.* Chicago: The Plymouth Press, 1925. 75 pp.

Osburn, W. J. *Corrective Arithmetic.* Boston : Houghton Mifflin Company, 1924. 182 pp.

Osburn, W. J. "Practice in the Fundamentals of Arithmetic," *Journal of Educational Research,* 10:356-63 (Dec., 1924).

Thorndike, E. L. "A Study of Errors in Reading," *Elementary School Journal,* 18:98-114 (October, 1917).

CHAPTER IX

REACHING THE INDIVIDUAL

In ancient and medieval times the pedagogue or tutor taught one pupil at a time. In more modern days, with the demand of the common people for education, the class became the unit of instruction. The trend is now again toward the individual. There is an effort to fuse the two apparently divergent ideas of individual and class teaching. It is recognized that true teaching is individual teaching. But the modern teacher, although still the teacher of one, must also be the teacher of many. I propose to describe in this chapter some of the methods and materials which have made this possible.

THE SURVEY TEST

The test movement has supplied many helps formerly unknown in teaching. One type of test is the survey test. It is designed primarily to give a bird's-eye view of a group. It is short and depends for its validity upon sampling the information or skill in a given field.

The status of the class. Only in a limited way can one base individual teaching upon the survey test. One thing, however, such a test does and does well—it indicates the general level of attainment of the class. The teacher learns whether he has a good, medium, or poor class. Moreover, a survey test given at the beginning of the year may be repeated at the end of the year. In the intervening months the teacher may have made a systematic attempt to apply certain methods. Special attention may have been given to individuals or to small groups in the class. The teacher will want to know whether or not his efforts have been rewarded. The best way to find this out is to give a survey test at the beginning and end of the period during which he wishes to measure progress.

Another use of the survey test. The teacher must often emphasize the elements of a totality. In English composition these elements may be sentence structure, comma-faults, punctuation, and the like. In handwriting they may be uniformity of slant, line quality, spacing, alignment, and letter formation. We divide in order that we may conquer, but there comes a time when the success of our teaching is to be determined by a general result. In short, we must find out whether after having divided we have succeeded in conquering. This is the time to use the survey test. In composition or handwriting the outcome of instruction even of the most detailed character is to be sought in the general merit of the production and not in the merit of an isolated element of it. At each step the practical teacher emphasizes definite elements. These he selects according to the needs of the pupils, but we do not expect pupils to refrain from expressing themselves until we have set up one by one these subordinate skills. In composition we may require a weekly or monthly theme; and, while some phase of the craftsmanship of writing may at any moment be to the fore, the pupil is always expected to turn in a total product. It is by this

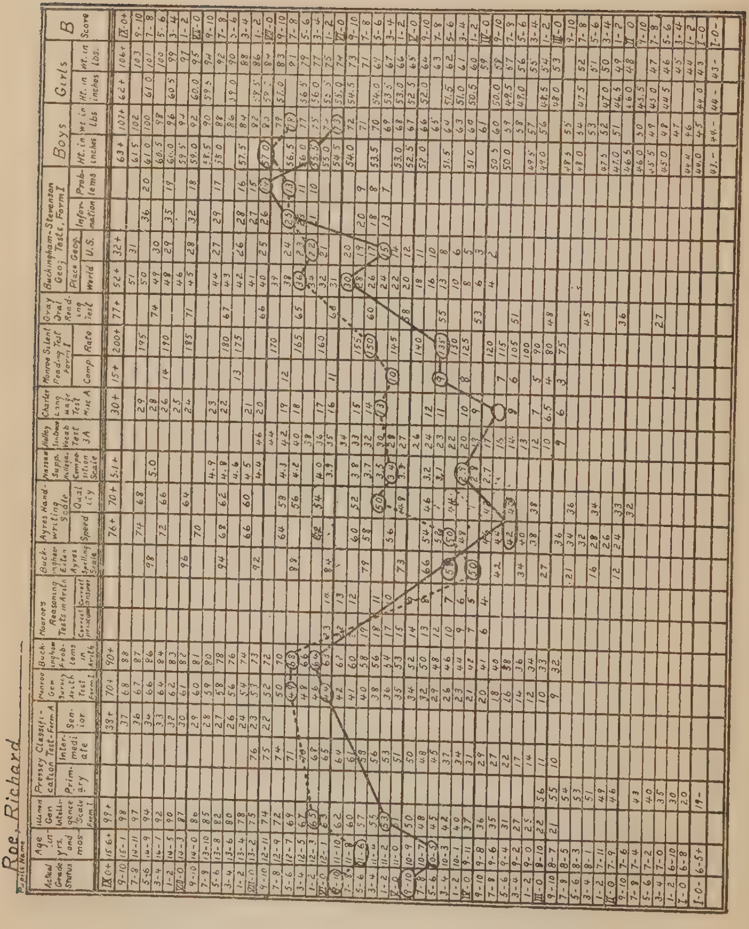

FIGURE 8.

total product, that we estimate the proficiency of the pupil and the success of our teaching. We have need, therefore, for a survey test.

INDIVIDUAL GRAPH CARDS

The standardized survey tests and scales have grade norms, and this fact has been utilized to devise card forms on which may be shown each pupil's success in as many standardized tests as he may have taken. The record is shown by a curve or graph. A card used for this purpose and in this way may be called an individual graph card. Such cards have been devised by a number of workers—for example, Courtis, Brooks, Arthur, and Goodrich. It seems to me that the best of them is the "Van Scoyck-Young Permanent Test Record Card" recently described by Brooks[1] and shown in Figure 8. It resembles Brooks' own card[2] but it is more complete.

This card requires no calculation whatever. The grade norms are printed directly on the card and are thus transferred from the test paper to the card without intermediate operation. In Figure 8 the ring around "9-10" indicates that at the time of testing the child in question was in the ninth or tenth month of the fourth grade. In the next column the ring around "10-5" means that he was ten years, five months of age. In the next column the circle around "53" shows that the pupil scored 53 on the Illinois General Intelligence Scale. The number 53, however, and the series of numbers in the same column are placed opposite the grade and the months for which they are standard. In this case the pupil's score on the Illinois General Intelligence Scale is shown to be equivalent to that of a child in the first or second month of the fifth grade.

---

[1] Brooks, S. S. "The Graphic Interpretation of Test Scores," *Educational Research Bulletin*, 4:366-71 (November 18, 1925).

[2] The Brooks individual graph record card published by the J. L. Hammett Company, Cambridge, Massachusetts.

Thus by encircling a pupil's scores and by joining these circles one may represent his performance in all his tests. The solid line in Figure 8 records the fact that Richard Roe was in the first two months of the fifth grade in respect to general intelligence, that he was at the beginning of the sixth grade in the Monroe General Survey Test in Arithmetic, and that he was of about the same status in the Buckingham Problem Test in Arithmetic. In handwriting, he scored at about the middle of the third grade in both speed and quality. In composition he was near the beginning of the fourth grade; in language, in the high third grade, etc. The "B" scores represented in the last column are the grade transversions of the raw test scores. In Chapter IV, I have called these "grade scores."

In Figure 8 the dotted line shows the performance of the same pupil a year later. This was when he was in the 9th or 10th month of the fifth grade. (See Column 1.)

DIAGNOSTIC TESTS

If a survey test shows that a class is weak in a given subject and indicates with some precision the particular pupils who are weak, we shall at once want to know what these weaknesses are. If a child fails to answer questions about a paragraph he has read, we shall want to know whether it was because of lack of vocabulary, failure to understand what he was to do, lack of control of the mechanism of reading, or lack of general intelligence. Again, if a child makes a poor showing in column addition, it will help us in applying remedies to know whether it was because he did not know the basic addition facts, because he was unable to "bridge the tens," because he had trouble with carrying, or because his attention span was not extensive enough to carry him all the way up the column.

Hence arise diagnostic tests—tests designed to give detailed information, tests intended to penetrate into the causes of general shortcomings. A testing program should involve the giving of one or more diagnostic tests immediately after a survey test has been administered.

INVENTORY TESTS

An inventory test is one which is designed to find out at the beginning of a proposed period of instruction what the pupils know and do not know concerning the material about to be taught. An inventory test will tell the teacher where to begin with each individual in respect to the particular instructional unit. Certain knowledge may be prerequisite. Certain facts of English grammar are needed before Latin can be studied. The one hundred basic facts in addition must be known before column addition can be learned. The inventory test may investigate the pupil's possession of prerequisite knowledge.

Pupils often know more than we think. The inventory test is concerned quite as much with finding out what a child knows as it is with finding out what he does not know. Students and pupils often know more about the work we are going to teach than we imagine. In a certain city school system the sixth-grade pupils at the beginning of the year were given ten questions on the course of study in geography which they were about to study. The following were some of the results: (1) In each class some of the questions were answered to the entire satisfaction of all the teachers. (2) Each question was answered satisfactorily by at least some of the pupils. (3) More than half of the pupils answered at least some of the questions satisfactorily.

With further reference to this point, it is represented by an instructor in educational psychology that when he gave to a class at the beginning of the term a test which he had previously used as a final test in the course, twenty per cent of the students passed the test.

In spelling it is admitted that some children know all the words and that all know some of the words of each assignment. One spelling book makes effective use of this fact. Children are tested on the next assignment before it is made.

It is undoubtedly a fact that we have underestimated the educative influences outside of the school. We are prone to believe that the only way a child can learn anything is through our ministrations. This is one of the reasons why we hold so tenaciously, and in the face of contrary evidence, to the idea that if a pupil is skipped there will be wide gaps in his knowledge from which he will suffer for the rest of his life.

How to make an inventory test. An inventory test may be a diagnostic test which has been printed for general use. A particularly effective type of inventory test, however, is the one made for local use. In this connection I suggest that you use some of the new-type techniques. The following steps may be observed: (1) Make a short list of the most essential prerequisite items. (2) Make a list of all the important items which you propose to cover in the period to which the inventory test relates. (3) Convert the items of both lists into test questions, using one or more of the new-type techniques. (4) Administer the test on mimeographed sheets which provide a definite space in which the pupil is to make his response to each question. (5) Analyze the results.

THE DOUBLE-ENTRY SCORE SHEET

The fifth step in this suggested procedure calls for an analysis of results. One way of doing this is to employ a double-entry score sheet. This device may be used to great advantage with every type of test or examination. Since, however, it applies admirably in this connection I shall describe it here.

The problem is to secure the most helpful information from the answer papers. The double-entry score sheet assists in doing this by bringing to view on a single sheet of paper all the errors and all the successes which every child in the class has made. Only one thing could be more complete and that would be an analysis of the mistakes to determine their nature. This the double-entry score sheet does not attempt to furnish. If an answer is wrong, it calls it wrong without reference to the reason. In a later section I shall take up the question of analyzing the errors on the pupils' papers. At the moment I am concerned only with the double-entry score sheet.

The term "double-entry" refers to the fact that there are both line and column headings on the score sheet. The line headings are the pupils' names, initials or numbers. The column headings are the test questions.[3] For the sake of brevity these questions may be entered by numbers or by brief jottings, which will call to mind the topic to which each one refers.

A fifth-grade teacher may wish to give a test on the physical features of the United States. McCall gives a set of true-false statements on this general topic.[4] Ten of these statements (which the pupil is to mark as true or false) follow. I have indicated by italics whether the statement is true or false.

1. In general the mountain ranges run east and west .... True *False*
2. Most of the rivers flow north .... True *False*
3. The Rocky Mountains are nearer the Pacific Ocean than the Appalachian Mountains ... *True* False
4. The Blue Ridge is in the Rocky Mountains .. True *False*
5. There are more active volcanoes in the west than in the east .... *True* False
6. The Mississippi River flows into the Great Lakes .... True *False*
7. All the following are tributaries of the Mississippi River: Arkansas, Missouri, Ohio ...... *True* False
8. The Big Sandy is the biggest river in the United States .... True *False*

---

[3] The placement of line and column headings may be reversed—that is, there is no virtue in having the pupils' names as line headings any more than as column headings. The number of pupils in relation to the number of questions and these in relation to the dimensions of the paper will determine the question. Put the greater number of headings along the longer dimension of the paper.

[4] McCall, William A. "A New Kind of School Examination," *Journal of Educational Research,* 1:33-46 (January, 1920).

9. It is easier to sink while swimming in the largest lake east than in the largest west of the Mississippi .... *True* False
10. The central portion of the United States is on the whole more level than the eastern or western portion .... *True* False

Of course, a true-false test should contain more than ten items, but these will suffice for purposes of illustration.

Figure 9 shows a double-entry score sheet for a class of twenty-eight pupils, to which the above test has been given. The top line reads, "Paul got the first, fifth, and eighth questions wrong." In other words, he did not know the general trend of the mountain ranges, nor the location of active volcanoes, nor the size of the Big Sandy River in relation to the other rivers of the United States. He made errors on three questions which, according to the right-minus-wrong method would give him a score of four. This method of scoring which is peculiar to the alternative-response test need not concern us in connection with the double-entry score sheet. The second line reads, "John got the fourth, seventh, ninth, and tenth questions wrong." The readings of the several lines will give the teacher information as to the knowledge and lack of knowledge of each pupil.

The columns are, however, just as important as the lines. The score sheet shows that the first question was answered wrong by ten pupils in the class, and, furthermore, it specifies the particular pupils who got it wrong—Paul, Henry, Frederick, etc. It is the same with all the other questions. The Rocky Mountains and the Mississippi River seem to be pretty well in hand. All but two of the ten pupils agreed that the Rocky Mountains were nearer the Pacific Ocean

DOUBLE ENTRY SCORE SHEET

City........................ School........................ Class.....

Teacher........................... Date.................. Class Score.....

Pupils	Questions										Total
	1	2	3	4	5	6	7	8	9	10	
Paul	x	..	..	..	x	..	..	x	..	..	3
John	..	..	..	x	..	..	x	..	x	x	4
Henry	x	x	..	x	x	..	..	..	..	..	4
Robert	..	..	x	x	..	..	x	..	x	x	5
James	..	..	..	..	..	..	..	..	x	..	1
Frederick ...	x	..	..	x	..	..	x	..	x	..	4
Albert	..	x	..	..	x	..	..	..	x	..	3
William	..	..	..	x	..	..	x	..	x	..	3
Charles	x	..	..	x	x	..	x	x	x	..	6
Raymond ...	..	x	..	x	x	..	..	x	..	..	4
Alfred	..	..	..	..	..	..	x	x	x	..	3
Thomas	..	..	..	..	..	..	..	..	x	x	2
Ralph	x	x	..	x	x	..	..	x	x	x	7
Carl	..	..	..	..	..	..	..	..	x	..	1
Mary	..	..	..	..	..	..	..	..	x	..	1
Edith	x	..	x	x	..	..	x	x	x	..	6
Catherine ..	..	..	..	..	..	..	..	..	x	x	2
Helen	x	..	..	x	x	..	x	..	x	..	5
Sarah	..	..	..	x	x	..	x	x	x	..	5
Ruth	..	..	..	..	..	..	..	..	..	..	0
Emma	x	x	..	x	x	..	x	..	x	x	7
Margaret ...	..	..	..	..	..	..	x	x	x	x	4
Bertha	x	x	..	x	..	..	x	x	x	x	7
Josephine ...	..	..	..	x	x	..	x	..	x	x	5
Gertrude ...	x	..	..	x	x	..	..	..	x	x	5
Blanche	..	..	..	x	x	..	x	x	x	x	6
Edna	..	..	..	x	..	..	x	..	x	..	3
Esther	..	..	..	x	..	..	x	..	x	..	3
Total	10	6	2	18	12	0	16	10	24	11	109

FIGURE 9.

than the Appalachian Mountains. None of the children thought the Mississippi River flowed into the Great Lakes. On the other hand, the ninth question was a poser. The purpose is to test whether the pupil knows that the largest lake west of the Mississippi contains more salt than the largest lake east of the Mississippi. McCall rightly criticizes this question when he says, "instead of measuring this, we may be testing whether a pupil knows that it is easier to sink in fresh water than in salt water." The question is a good example of complex wording and indirect statement. Only rarely should such questions be used. The class was not very familiar with the Blue Ridge Mountains (question four) or with the tributaries of the Mississippi (question seven).

Enough has been said, I think, to show the value of the double-entry score sheet. Up to a certain point it permits the study of individual errors as well as individual successes. It shows the teacher in a rather detailed way—although not in the most detailed way—where to place emphasis in general teaching and what pupils to select for individual instruction. With respect to these pupils it likewise indicates with some precision what the individual instruction should be.

MORE DETAILED DIAGNOSIS

The most valuable information obtainable from testing is usually neglected. Teachers give tests—no doubt you yourself have given tests—and after they have served the needs of the moment they are either thrown away immediately or filed and forgotten. This procedure applies just as much to informal testing as to the more formal use of standardized materials. Not half the use is made of the responses of pupils to these tests as could and should be made. I propose to show one way in which test data may be utilized. I suggest this as an opportunity open to every teacher for instructional research.

Not long ago a certain superintendent gave three tests. These were: the Buckingham Problem Test in Arithmetic, the Monroe General Survey Test in Arithmetic (a test of abstract number work), and the Stone Narrative Reading Scale. The results of this testing were duly tabulated. Then the superintendent realized that there seemed to be little else for him to do. Accordingly, he requested some help in taking the next step, that is, in modifying instruction in the light of the test results. I was fortunate enough to be consulted in this matter and as a result of my study I became convinced that a rather definite procedure is possible in analyzing test results.

A METHOD OF ANALYSIS

General status. In the first place, we shall want to know in general where each class stands in the abilities measured by the tests and something about the rank and distribution of scores. This is the kind of information usually secured, and this is where the analysis generally stops.

The child as the center of thought. In the second place, the center of our thinking from this point on should be *the child and not the test.* This necessitates not only a different attitude but a different procedure. For example, we must actually arrange the papers *by pupils* not by tests—all the papers of the first pupil, then all the papers of the second pupil, etc. When this is done we can begin to study everything about each child. The errors on the papers will give us our best information. Each paper, so far as it presents errors, is a clinical picture. Each error has a meaning and a peculiar causation. If we can find these, the individualizing of instruction will not be difficult.

Arrangement by pupils. In the third place, we must arrange our work systematically. For example, in the city to which I have referred, I had the papers for each class assembled so that the three papers for each child were together. Then I had these individual packages arranged in the order of the scores on the *problem* test. On top of the pile for any given class I had the three tests of the child who got the highest score in problems; next were the papers of the child whose problem score was next high; and so on.

In this case, therefore, each child's problem paper becomes the point of departure. We have his rank on problems. Then we may examine his abstract paper and his reading paper to see what bearing his errors on these papers have upon his errors in problems. Of course, the papers could have been arranged according to size of score in either of the other tests. From the nature of the case, however, we are more interested in the bearing that reading has on ability to solve problems than we are in the bearing that problem solving has on reading. Similarly, the influence of computing ability on problem solving is more important than the reverse influence.

The total score may be misleading. In the fourth place, it is the papers, item by item, and not the total score on each paper, that must be examined. If a child is low in Monroe's test and if we note only the score, we have no guidance for remedial work except perhaps to drill him on every type of example represented on the test. This would indeed be coarse work. Again, if we regard total scores only, we shall often delude ourselves into supposing that all is well. One of the pupils in this eighth-grade class was somewhat above standard in the Monroe test. Yet he was entirely unable to add, subtract, or divide fractions; and he gave no evidence of knowing how to handle the decimal point in division. His high score was secured through unusually good work with integers.

The interview. In the fifth place, it will sometimes be necessary to supplement the examination of the pupil's paper by an interview with him. While there is much to be learned from what the pupil sets down, and while the examination of written errors and the determination of their causes is fascinating, the pupil frequently leaves out the evidence of what he did. Consider the following problem:

> George walks at the rate of 2 1/3 miles per hour. Henry starts with him and walks in the same direction at the rate of 3 miles an hour. How many miles apart will they be in three hours?

In response the pupil writes "2½ miles" and nothing more. The answer is wrong, but why it is wrong we cannot tell.

Accordingly, you should interview the pupils whose papers do not give you the evidence you need. Have them answer the test questions orally and tell you the reasons. You will be surprised at the result. In problem work the cues which lead children to add, subtract, multiply, or divide are often very curious. For example, if it says "times" in any connection, you multiply. The problem may be: "Joseph has twenty-five cents, and it costs five cents to ride on the merry-go-round. How many times can he ride?" Again, some children will tell you that if there are several numbers in the problem, you add; and that if there are only two numbers, you multiply—or perhaps divide.

THE METHOD ILLUSTRATED

I shall now illustrate the kind of analysis I have been describing by using the papers of an eighth-grade class in the city to which I have referred. The papers have been arranged in the manner I have suggested. I shall give special attention to the poorest papers—this for the sufficient reason that it is the errors rather than the successes that are illuminating. The teacher, however, should examine all the papers. The good papers will, of course, require little time.

A capable pupil. Before looking at some of the poorest papers, however, I should like to refer to the abstract arithmetic paper which had the highest score. The score was 101, which is so high that we have no grade norm by which to represent it. If we extend the grades upward, it will correspond to the first year of college; and no doubt this eighth-grade pupil would in truth equal college ability in work called for by this test. He was both rapid and accurate. He had only five errors on his paper.

Three of this pupil's five errors were in column addition and two were in subtraction of fractions. In the latter case he overlooked the minus sign between the fractions and correctly added them. So far, therefore, as the handling of numbers is concerned all his errors were in addition. Two of them were due to a peculiar cause—a cause which would only become evident upon a close scrutiny. First, I observed that both these errors were in unit's place. Second, but only after a considerable interval, I noted that both his additions were correct if the unit digit of the top addend were omitted. By an inference which presents every evidence of being satisfactory, I concluded that this pupil adds downward and that in his skipping from the bottom of the previous example, where he had been writing the answers, to the top of the next example, he had overestimated the speed at which he could pick up the next task. Apparently, he slipped by the first figure and fixated the second one.

We hear much about remedial instruction—and rightly so. But the real trouble is to find out what is the matter. If we know the causes of error we very frequently know what to prescribe. In the case of this pupil no remedial instruction is necessary. Almost the entire case is covered by a single item and that item can be taken care of in all likelihood by bringing it to his attention. This may serve to illustrate the fact that remedial work does not always have to resort to practice material or formal instruction.

Individual differences. The range of talent in this class was very wide. In reading, the rate was from something like fourth-grade performance to a performance three times as great as that of the average eighth-grade pupil. In comprehension of reading, the range was apparently quite as wide. In problem solving, the poorest pupil showed about sixth-grade ability and the best about tenth-grade ability. In abstract arithmetic the poorest performance would be surpassed by most fifth-grade children; while the best, as I have already indicated, might be represented as equal to that of a college freshman.

It is clear, therefore, that before the teacher of this class can successfully "reach the individual" he must have the assistance of his superiors in the matter of regrading. As I looked over these papers I wondered whether the teacher really knew the significance of these ranges. Very likely they mean that no remedial work which this teacher can do will serve to bring some of these pupils "up to grade." Individual instruction with the poorest of these pupils would be much better handled in the seventh grade.

Frances' papers. As I began to examine the papers of the least capable pupils, I first became acquainted with Frances. She got the first problem wrong. This was the problem:

> A man bought a house for $7,250. After spending $321.50 for repairs he sold it for $9,125. How much did he gain?

Frances reasoned correctly, but she made a mistake in subtracting the sum of the items of cost from the selling price. She came to grief in taking 7 from 12, which she found to be 4. From the fact that she put a "1" above the "2" in the minuend, it is clear that she arrived at the result not because of a belief that 7 from 12 leaves 4, but because she "borrowed" 1 from 2, although borrowing was not the correct procedure. Frances, therefore, failed on this problem because she made a mistake in computation. Moreover, we can go further and say precisely what the nature of this mistake was.

The next problem on which Frances failed was the following:

> If 0.78 of the weight of potatoes is water, how many pounds of water are there in a bushel of potatoes? (A bushel of potatoes weighs 60 pounds.)

Again the reasoning was correct. Sixty was multiplied by 0.78, the example being set down like this:

$$\begin{array}{r} 60 \\ .78 \\ \hline 480 \\ 430 \\ \hline 47.80 \end{array}$$

The wrong answer was obtained because in multiplying 6 by 7 she got 43. At this point I turned to Frances' paper in abstract arithmetic. Test No. 2 of the Monroe series is on multiplication. On this test Frances had three occasions to multiply 6 by 7. In every instance she got 43, just as she did in the problem test. Here again the scrutiny reveals facts which are in themselves sufficient to indicate the remedy. This pupil consistently fails on a certain combination. Our task is therefore clearly indicated. No elaborate procedure is necessary.

Returning now to the problem test, we note that Frances got the next problem wrong and that she did so for an entirely different reason. In this problem two boys agree to mow a lawn for ninety cents. One works four hours and the other one hour. The pupil is required to distribute the money. Frances multiplied 90 by 4 and gave the boy who worked 4 hours, $3.60. Then, not to leave his partner penniless, she gave him 90 cents—although her basis for this is not entirely evident. The important point is that she completely disregarded the agreement on the part of the boys to do the whole job for 90 cents. Evidently we have here a reading difficulty, and this suspicion is strengthened by an examination of her reading test. Whereas the typical eighth-grade child should score 11 or 12 in comprehension, she scored only 5. The remedy, therefore, for her failure to understand the conditions of the problem lies in the field of reading. She should receive the benefit of instruction and drill designed to improve her comprehension of printed material. Frances got two more problems wrong for the same reason—failure to read and understand the conditions imposed.

Remedies for difficulty in reading problems. As to the remedial instruction which may be employed to meet this difficulty of the defective reading of problems, I wish to emphasize the fact that we must train for specific abilities if we wish to secure them. In other words, we cannot rely upon the transfer of training.

We are learning with increasing certainty that the abilities of children are specialized to a remarkable degree. We must specialize our treatment correspondingly. If the difficulty is a lack of ability to read arithmetic problems understandingly, we must not merely lengthen the reading period. We certainly should not seek a remedy by subjecting Frances and her classmates to more prolonged oral reading. Indeed, it is precisely this treatment already unduly prolonged which has contributed to the difficulty. Moreover, even if we are teaching silent reading according to approved methods, we shall secure—so runs the evidence—far greater improvement in problem solving if we use *problems* for reading purposes. It is even arguable that after the mechanics of reading have been taught, the reading period might well be abandoned in favor of teaching reading in the arithmetic, geography, history, and science periods. It has been contended that there is no more reason why reading should be taught in connection with literature than that it should be taught in connection with any other school subject.

There are various ways in which we may train Frances and those who have this reading difficulty. We may, for example, employ some of the new-type techniques of test making, such as multiple choice.

Here, for example, is a problem which Frances evidently failed to read correctly:

> A real estate dealer made a profit of $5,000 on a plot of ground. The rate of profit was 20 per cent. Find the cost of the plot.

To a child who senses the situation it is immediately evident that this fortunate realtor who made a profit of $5,000 must have handled a big deal. Frances, however, got no such notion. Her answer was $80.00 as the cost of the plot of ground. Of course it is possible for a dealer to buy for $80.00 and sell at a profit of $5,000, but it would be quite impossible to represent this as a profit of 20 per cent.

Suppose this problem, together with other problems, were presented in a test designed to measure accuracy in reading this class of material. We might then ask the reader to indicate which of the following were *given:* (a) the total cost, (b) the total selling price, (c) the profit, (d) the cost per acre, (e) the rate of gain, and (f) the selling price per acre.

It is likewise suggested that the *thing required* in a problem may also be handled by the same technique. It is also possible to use the same method in requiring from the pupil the operation or operations to be performed. It is not practicable to do this for problems involving more than two operations. It is surprising, however, how rich a variety of problems may be presented without requiring more than two operations.

With further reference to training and testing pupils as to the content of problems, questions may be asked which have to do neither with the essential data nor with the requirement. For example, Wilson[5] describes the automobile trip of the Brown family. Certain problems are to be solved, but some questions are asked which have no essential bearing on the solution of the problems. The pupil is asked how many different prices they paid for lodging and breakfast, although the total cost is all that the problem on this point requires. Again, the pupil is asked what different things Mr. Brown had to consider in figuring the cost of running the car. The problem on this point gives the actual cost for gasoline, oil, etc., and requires that the total cost be computed. Such questions as these undoubtedly sharpen the attention of the pupils to the content of problems.

---

[5] Wilson, Estaline. "Improving the Ability to Read Arithmetic Problems," *Elementary School Journal,* 22:380-6 (January, 1922).

A suggestion as to absurd answers. Frances got a number of absurd answers to problems. This is a very common type of error. One remedial device for this sort of trouble is to practice and test pupils on problems in which no solutions are required but in which several answers are presented. The requirement is that without solving the problem they shall determine which of the given answers is *reasonable*. Observe that none of the answers needs to be precisely correct. One of the problems which Frances got wrong gives the distance from New York to Chicago together with the schedule time. Due to floods the train goes only 296 miles in the first seven hours. The pupil is required to find how many miles per hour it must go to reach Chicago on time. If we were using this problem in our remedial work we might set up the following series of answers from which the one reasonable answer was to be selected: 1,000, 2 1/3, 91, 46, 24.75.

Harriet's papers. Without attempting to exhaust the possibilities of instructive inferences from Frances' papers, we may pass to those of Harriet, who shared with Frances the poorest position in the class in respect to problem solving. One begins to find, even upon consideration of this second paper, a duplication of errors. For example, Harriet, like Frances, failed to solve the first problem. Like her, she reasoned correctly, adding the two cost items and subtracting their sum from the selling price. Like Frances, she obtained the correct sum, but made a mistake in subtracting it from the selling price. Nor does the similarity stop here. She actually made the same numerical error. You will recall that Frances subtracted 7 from 12 and got 4. Harriet did the same. She did not, however, put a "1" over the "2" in the minuend as Frances did. We cannot therefore be sure whether in Harriet's case the difficulty was an instability of the 12 less 7 combination or a confusion as to when the process of borrowing applies. A few moments' interview with her would clear up this point. Examination of her test in abstract work yields no conclusive evidence.

Harriet had a greater repertoire of mistakes in problem solving than Frances. To each of the two boys who agreed to cut a lawn for 90 cents she gave half the money, ignoring the fact that the first boy worked one hour and the second four. She copied a number wrong in the example about the amount of water in a bushel of potatoes. She made a mistake in finding the total of a bill of goods. She made the belated railroad train develop the astonishing speed of 179 196/351 miles per hour. In this latter case, instead of dividing the distance by the time, she divided the time by the distance. She added an indefinite number of zeros to make the dividend "contain" the divisor, and her industry in thus pursuing the division seemed to be limited only by the space at her disposal. The lucky real estate dealer with his 20 per cent profit of $5,000 is made to purchase his plot of ground for $4,000.

Remedial measures in the case of this pupil may evidently be much the same as those in the case of Frances. Reading difficulties, misapprehension of conditions, acceptance of absurd answers—these are fundamental in both cases. Her general inability to translate printed material into thought is shown by her reading test. After reading the passage in the Stone test, she was able to answer only 9 out of 20 questions based upon it.

An examination of Harriet's paper in abstract arithmetic is interesting. In the first place, her total score was above the grade norm. The temptation in such an instance is to pass over the pupil's paper with the easy assumption that conditions are satisfactory. In the case of Harriet, however, nothing could be farther from the fact. She scored only 4 in addition. A critical examination of the way in which her errors were made discloses the fact that half of them occurred in adding the unit columns. In other words, there is a decided tendency on her part to make a mistake at the beginning of an example. It is not difficult to see how this comes about. Her difficulty is analogous to that of the very capable student already mentioned. The vigor of her attack upon each new example tended to outrun her ability.

In division she was decidedly poor. She tried 10 of the examples and got 5 of them wrong. Yet all these failures were due to a single recurring cause. She was not able to estimate the number of times the divisor was contained in the significant figures of the dividend or in the remainder. Moreover, when she estimated wrongly, she was quite without resources. For example, having estimated that 79 would be contained in 368 five times and having multiplied 79 by 5 only to find that the product exceeded 368, she was entirely unable to draw the inference that the correct figure would probably be 4 rather than 5. Each time she failed to make the correct estimate, she abandoned the example as soon as it was evident that her estimate was wrong.

To mention this sort of difficulty is, as in so many other instances, to suggest the remedy. What Harriett needs is drill in estimating the number of times a number of two digits is contained in a given dividend. She also needs drill on self-correction when the first estimate is wrong.

When it came to fractions, Harriet might subscribe to at least a portion of the ditty, "The rule of three perplexes me and fractions drive me mad." She is totally unable to add or subtract two fractions. Her method is unique. She adds or subtracts the numerators according to the sign between the fractions. If the sign is minus, she subtracts either the first numerator from the second or the second from the first with entire impartiality. As to the denominators, she multiplies them for the denominator of the answer; but she has a short cut which affects the treatment of both the numerators and the denominators. She cancels as in multiplying. If one were to put into the form of a rule the process she employs, the wording would be somewhat like this: "Cancel if you can; then add or subtract the numerators, and multiply the denominators."

Test No. 6 of the abstract arithmetic test combines multiplication and division of fractions. Harriet saw no difference. To her everything was multiplication. The result was that, barring accident, she was right in multiplication and wrong in division. One of the accidents which interfered with her securing correct answers to the multiplication examples was the fact that when the numerators cancelled completely into the denominators, she represented the answer as a whole number instead of as a fraction with "1" for the numerator.

Finally, in the test on decimal fractions Harriet showed no clear indication of an ability to handle the decimal point.

Thus it will appear that a pupil may have an entirely satisfactory total score from the point of view of the grade norm, yet fail completely on certain parts of the test.

Esther's papers. I am going to allude briefly to the third set of papers. Esther is of the type "slow but sure." She is not a very good example of the type because she isn't very sure. She is, however, slow; and in spots she is sure. She does a few of the first problems correctly—they are arranged in order of difficulty—and then gets them all wrong without exception. In the latter part of the test she alternates between wrong principle and absurd answers as the chief source of difficulty.

In the abstract test she does pretty well on integers but she does not try many examples. Fractions, however, drive her quite as mad as they do Harriet. Her method in addition and subtraction is much simpler than Harriet's. If it says add, why, you add consistently. You add the two numerators for the numerator of the answer and the two denominators for the denominator of the answer. If it says subtract, you carry out the direction with equal thoroughness. As in the case of Harriet, it appears to make no difference to Esther which way the subtraction is operated. In multiplication and division of fractions, she tried but three examples and got them all wrong. The fraction tests are evidently the spots where, though undeniably slow, she is far from sure.

As to individualizing instruction in Esther's case, two things stand out prominently. The first is that like others in the class she needs to be taught fractions from the beginning. This, of course, can be done in groups—perhaps with the class as a whole. In the second place, Esther should be given practice material under timed conditions. Very likely she is constitutionally of slow reaction, but it is also likely that the limit of her reaction time is nowhere nearly reached in her performance on these tests. She needs the stimulus of the timed control. Like the other members of the class—and here the teacher can also work with groups—she needs to be taught to *read* problems.

We need not prolong further the examination of these papers. If you make an error analysis of a set of papers written by your pupils, you will be surprised to note as you proceed how the different types of error prevalent in the class begin to emerge. It is true there will be difficulties peculiar to each child, but there will be other difficulties which several children exhibit, and perhaps some which are common to the entire class.

In many instances the mere knowledge of the existence of the trouble will suggest the remedy. We should teach, drill, and test the pupils in the precise activity which constitutes the difficulty. Our main trouble is one of precision. We tend to go shooting with a blunderbuss and thus to waste our ammunition. The effort to raise ability in arithmetic by giving a generalized treatment is a weak and wasteful effort compared to that which is directed to the strengthening of specific weaknesses and the elimination of particular disabilities. Our individual work can, it is true, be conducted in groups where group difficulties exist; but much of it will have to be on an individual basis. Nor can that individual basis be secured except through trial of pupils, an analysis of their failures, and an application of teaching, drill, and further testing on types of material accurately adjusted to the disabilities which the initial testing reveals.

PRACTICE MATERIAL

I have tried to make it evident that reaching the individual depends upon finding out the condition of the individual. This means testing. It also means intelligent and persistent utilization of the results of testing.

One way to go about this is to analyze into subordinate elements the subject-matter which we are attempting to teach. These elements will be most valuable when they correspond to children's difficulties. Thus the topic "addition of whole numbers" may be divided into steps, each corresponding to a new difficulty as the child masters the subject. The following steps have been distinguished:

COMBINATIONS

(1) The fundamental number facts whose sums are less than ten. These are called the easy combinations.
(2) The fundamental number facts whose sums are ten or more. These are called the hard combinations.
(3) Higher decade combinations without bridging the tens, *e.g.*, $25 + 2$, $31 + 8$.
(4) Higher decade combinations with bridging the tens, *e.g.*, $13 + 8$, $25 + 7$, $36 + 5$.
(5) Two two-place addends without carrying, *e.g.*, $\begin{array}{r}13\\ \underline{24}\end{array}$

COLUMN ADDITION

(Downward addition is assumed).

(6) Three one-place numbers whose sums are less than ten, *e.g.*, 2
3
4
—

This introduces for the first time the *form* of column addition. Two successive number facts are employed and one of them is "unseen"—that is, in one case an addend is not visible. All column addition involves these two points of difficulty—namely, successive use of more than one combination and the fact that all combinations after the first are unseen. Since in this step the sum is restricted to less than ten, both the combinations will be easy combinations.

(7) Three one-place numbers, one combination being "easy" and the other "hard," *e.g.*, 2
3
7
—

(8) Three one-place numbers, one combination being hard and the other a higher decade combination without bridging, *e.g.*, 7
6
2
—

(9) Three one-place numbers involving a hard combination and a higher decade combination with bridging, *e.g.*,
7
6
8
—

(10) Four one-place numbers.

(11) More than four one-place numbers. Steps ten and eleven are an approach to the difficulty of the attention span. It is hard for children to hold their attention to adding a long column.

(12) Two three-place numbers without carrying, *e.g.*, $\begin{array}{r}274\\503\\\hline\end{array}$ The new point here is merely the extension of addition to large numbers, including the reading and writing of such numbers.

(13) Two two-place numbers with carrying. The new difficulty is, of course, carrying.

(14) Three two-place numbers with carrying.

(15) More than three two-place numbers.

(16) Two three-place numbers with carrying. Three subdivisions of this step are possible: (a) when the carrying is from the units column only, *e.g.*, $\begin{array}{r}274\\419\\\hline\end{array}$; (b) when carrying is from the tens column only, *e.g.*, $\begin{array}{r}274\\562\\\hline\end{array}$; (c) when the carrying is from both the units and the tens column, *e.g.*, $\begin{array}{r}274\\568\\\hline\end{array}$

(17) More than two three-place numbers.

(18) Irregular columns, including one-, two-, three-, and four-place numbers and three or more addends.

In some of these steps one- and two-place numbers should occasionally be included, although the type may not call for them. Thus a one-place number should sometimes be given in steps fourteen and fifteen. Similarly, two-place numbers and one-place numbers may be included for variety in steps sixteen and seventeen.

Hillegas has made an analysis similar to the above for all four of the operations with integers.[6] Analyses of the same general type may be made in other subjects, although in none of them, so far as I know, has the analysis idea been carried as far as in arithmetic. Tyler has shown in a study not yet published the possibilities in Latin, and I myself have made a beginning in algebra.[7] When a school subject or topic has thus been analyzed into "unit skills," the use of the analysis for individual instruction requires, first, that an inventory test be given to find out the abilities and disabilities of the pupils. This test must cover the items into which the subject has been analyzed. For example, in connection with the above analysis of integers, after the weaknesses of each child have been determined, the teacher may then provide him with practice material corresponding to the step to which one of his difficulties belongs. This material he will use until he gives evidence of having mastered the difficulty in question. He will then take a test pertaining to this difficulty, and if he passes it he will go on to the practice material covering the next difficulty.

---

[6] Hillegas, M. B. *Teaching Number Fundamentals.* Philadelphia: J. B. Lippincott Company, 1925. 98 pp.

This practice material should be abundant and it should be provided with answers. Each type should be definitely concerned with a difficulty which enters into the general mastery of the subject. At any given time one child may be using one type of practice material and another, another. Moreover, each pupil will advance at his own rate. He may take one day or he may take two weeks to attain satisfactory skill in a particular item.

Since the preparation of practice material is in general too great a task for the teacher to assume, a thoroughgoing application of individual instruction through the use of practice exercises can proceed only when these exercises are furnished to the teacher in prepared form. This has been done in arithmetic by Courtis[8] and by Clapp.[9]

---

[7] Buckingham, B. R. "Suggestions for Teaching IX-B Algebra," *Educational Research Bulletin*, 4:243-50, 260, September 9, 1925.

Individual instruction at Winnetka. One entire school system in the country has been organized on the basis of individual instruction of the sort I have been describing. The schools at Winnetka, Illinois, under the superintendency of Dr. Carleton W. Washburne, have been administered for several years with the idea of providing individual instruction at least in the tool subjects. In the course of the development of this plan at Winnetka and as a necessary part of it, certain materials have been devised. Some of them have been issued in a preliminary form. In his preface to these materials Dr. Washburne has this to say:

> "While the materials themselves need much work before they will be perfected, the basic principles of individual instruction have gone beyond this experimental stage. No one longer doubts the wide variation that exists among children. No one doubts the desirability of fitting public schools to these differing individuals.
>
> The Winnetka Schools have proved the practicability of the three basic principles of individualizing the work in the common essentials: (1) goals of achievement specifically and definitely stated in place of the time unit; (2) complete diagnostic tests covering each of these goals; (3) self-instructive, self-corrective practice materials leading up to the tests.
>
> We hold these three principles to be universally applicable. Through their application individual instruction in the public schools becomes possible."

---

[8] Courtis, S. A. *Standard Practice Tests in Arithmetic.* Yonkers, New York: World Book Company, 1920.

[9] Clapp, F. L. *The Clapp Drill Book in Arithmetic.* Silver Burdett and Company.

Thus, the work at Winnetka, like all work in individual instruction, proceeds on the well established basis of individual differences. By including certain studies in the elementary school, society has in effect declared that the knowledges and skills pertaining to these subjects are needed by all children. But children differ greatly in ability to attain these ends. In the ordinary graded school we attempt to take account of this by putting all children over the same ground in the same length of time and by recognizing—none too explicitly, to be sure—that the product will differ. Dr. Washburne points out that instead of a difference in product, we should have a difference in the time—that all children should learn these socially necessary knowledges and skills to the same degree, but that they should be allowed longer or shorter times for doing this in accordance to their abilities. A fuller statement of the Winnetka plan may be obtained from the writings of Dr. Washburne.[10]

While it is true that the preparation of large amounts of practice material is beyond the resources of the teacher, it is nevertheless true that individual instruction through the use of informal material of this character is entirely practicable. I do not mean that you can do it for all the children in your class in all subjects. But you can do it for a few children in one or two subjects. The theory underlying the plan is sound and it ought to be applied whenever it can be applied. Moreover, the method is simple, It consists first, in testing to find out the weaknesses of the children; second, in providing materials by which they may *help themselves* to overcome these weaknesses; and third, in retesting the pupils to find out the extent to which the weaknesses have been overcome.

---

[10] For example, "The Individual System in Winnetka," *Elementary School Journal,* 21: 52-68, September, 1920. "Educational Measurement as a Key to Individual Instruction and Promotions," *Journal of Educational Research,* 5:195-206, March, 1922.

The experience of a third-grade teacher. An idea of the feasibility of reaching the individual under practical classroom conditions may be gathered from the experience of Miss Geiger, a third-grade teacher in Cincinnati.[11] She covered the first step (finding the individual weaknesses) by giving four standardized reading tests, each designed to test a different type of reading. In consequence she divided her class into three groups according to their needs as shown by the tests. She then adopted two lines of procedure: first, she used certain general methods for each group, and second, she applied certain individual methods in special cases.

Her object was to develop "intelligent comprehension and a standard rate." In her report she gives in detail some of the most successful remedial measures for group work. Ultimately, however, her problem resolved itself into a problem with individuals. She describes a few cases and tells what she did. I wish I might take the space to quote the case of the boy who loved books and read them in his slow plodding way and who, because of intelligent individual attention, added fifty or sixty words a minute to his reading rate. The story of the boy who flew at words that seemed to get no further than his eyes would be equally interesting. The teacher used entirely different methods with him. They must have been effective for he increased his comprehension of reading (November to May) one and a half years according to the Monroe test and more than two years according to the Courtis test. You will observe that the teacher gave the standardized reading tests near the end of the school year as well as at the beginning. Her class at no time had less than forty-five pupils, and she gave no more than the regular amount of class time to reading.

---

[11] Geiger, Ruth. "A Study in Reading Diagnosis," *Journal of Educational Research,* 8:283-300 (November, 1923).

SUMMARY

All teaching is individual teaching. After a period during which teachers aimed at the whole class—personified by the "average pupil"—the tendency is now toward the individual. "Reaching the individual in the group" is a touchstone for good teaching.

The survey test is indirectly useful in individual teaching because it affords a broad basis. It tells the general level of attainment of the class. Moreover, it affords a measure of the success of teaching specific topics. If handwriting is analyzed into elements and each element is taught, the success both of the analysis and of the teaching may be learned by applying a scale for general merit. Again, survey tests when grouped into "batteries" give the general status of the individual over a wide range of subject matter. In order to bring such information into systematic form, individual graph cards have been devised. They show at a glance the grade level of a pupil in each of several subjects.

The diagnostic test gives details. Such a test may be used, at the beginning of a teaching unit, as an inventory test. Teachers may likewise make their own inventory tests.

The purpose here is to find (1) what the child knows and does not need to be taught, and (2) what he does not know and needs to learn. It has been found that in some subjects pupils have considerable information before they receive formal instruction.

As an aid in analyzing the results of giving any test or examination I have suggested the double entry score sheet.

More detailed analysis, however, should be carried out by a study of the pupils' papers. This appears to me to be the most helpful way of reaching the individual on the basis of testing. I have described a method of making this analysis and have applied it in some detail to an eighth-grade class tested in problems, abstract arithmetic, and reading.

Although the analysis of pupils' errors will in many cases suggest informal remedial measures, individual teaching of a thoroughgoing sort requires the use of practice material. In order to learn what material is needed, a subject or large teaching unit is divided into smaller units. Tests and practice materials are then devised for each of these units. A diagnostic or inventory test discloses the units which the pupil has not mastered and the appropriate practice material is then given him. When he has apparently mastered the unit to which the practice material belongs he takes another test. If he passes it, he goes on to another unit.

This is individual teaching on a large scale. It is beyond the powers of the individual teacher to do it. It is an administrative matter. I have introduced it to show how far individual instruction may go.

But the teacher has large opportunities for individual teaching independent of any administrative organization. A knowledge of individual needs may be obtained, and such knowledge will quite generally suggest means of meeting the needs. I have taken from recent reports one by a third-grade teacher who, with a class of forty-five pupils, *reached the individual.*

BIBLIOGRAPHY

Arthur, Grace. "An Individual Record Card for Preserving Test Data," *School and Society,* 18:356-57 (September 22, 1923).

Brooks, Samuel S. *Improving Schools by Standardized Tests.* Boston: Houghton Mifflin Company, 1922. Chapters IV, V, and VI.

Brooks, Samuel S. "The Graphic Interpretation of Test Scores," *Educational Research Bulletin,* 4:366-71 (November 18, 1925).

Buckingham, B. R. "An Opportunity for Instructional Research," *Phi Delta Kappan,* 5:10-11, 13-14 (April, 1923).

Courtis, S. A. *Standard Practice Tests in Arithmetic.* Yonkers, New York: World Book Company, 1920.

Dalman, Murray A. "Hurdles, A Series of Calibrated Objective Tests in First-Year Algebra," *Journal of Educational Research,* 1:47-62 (January, 1920).

Geiger, Ruth. "A Study in Reading Diagnosis," *Journal of Educational Research,* 8:283-300 (November, 1923).

Goodrich, T. V. "Putting Test Scores in a Form to Facilitate Comparisons," *Journal of Educational Research,* 8:377-88 (December, 1923).

Kallom, Arthur W. "The Importance of Diagnosis in Educational Measurement," *Journal of Educational Psychology,* 10:1-12 (January, 1919).

CLAPP, F. L. "The Clapp Drill Book in Arithmetic," Silver Burdett and Co.

WASHBURNE, CARLETON W. "The Individual System in Winnetka," *Elementary School Journal,* 21:52-68 (September, 1920).

WASHBURNE, CARLETON W. "Educational Measurement as a Key to Individual Instruction and Promotions," *Journal of Educational Research,* 5:195-206 (March, 1922).

WILLING, MATTHEW H. "The Encouragement of Individual Instruction by Means of Standardized Tests," *Journal of Educational Research,* 1:193-8 (March, 1920).

CHAPTER X

THE TEACHER AS A RESEARCH WORKER

Here and there in the foregoing chapters I have taken occasion to point out the strategic position in which I believe the teacher is placed with reference to educational research. In this chapter I desire to develop this idea rather more fully. In so doing I wish to emphasize two points of view. The first is that education needs the teacher as a research worker. The second is that the teacher and the teacher's calling will be benefited by the teacher as a research worker.

Experiments in learning antedate the appearance of either educational or intelligence tests. Since learning is the obverse of teaching, the hand to which teaching is the glove, you might surmise that this long line of research in learning would have profoundly affected the practice of teaching. It has failed to do so to anything like the degree which its intrinsic importance would lead us to expect.

As long as learning experiments are handled by psychologists alone we shall make slow progress so far as education is concerned. Many psychologists are not even interested in education. Under these circumstances we have a lot of piecework but no quantity production. The only persons who can supply the need in this respect are the teachers.

Winch, who has for years been one of the foremost supporters of the application of scientific methods in the study of education, has aptly said in this connection: "I have on more than one occasion published the opinion that without the coöperation of teachers there will never be in any appreciable sense a science of education. . . . If teachers stand aside and do not help, they will only have themselves to thank if methods are imposed upon them from the results of experiments which they do not understand, which are not really applicable to their work, and in which they have no share."

It may surprise you to learn that most of the learning experiments have been based upon results from but a few persons. The psychologist in his laboratory does not have access as the teacher does to large numbers of learners. The work of the psychologist is of high value, but it is valuable, so far as education is concerned, largely because it opens up new fields and sets up new problems. These fields will not be thoroughly explored nor these problems adequately solved until the psychologist in the laboratory is supported by the teacher in the classroom.

You may be prepared to believe that you can experiment in reading, spelling, and arithmetic but you may not see how you can conduct learning experiments as such. There are two observations to make in this connection. One is that we need experimentation where reading, spelling, and arithmetic—as well as literature, science, history, and foreign languages—are the learning material. The second is that even with the materials hitherto used in the laboratory you can be of real service. I have lately been in touch with a number of research workers in foreign lands and I find that teachers' associations are considerably interested in problems which we think of in this country as distinctly psychological.

Some of the experiments performed by the psychologist can be conducted by you, and there are others, available to you, which he cannot easily manage. You will need some direction and at first you will make mistakes. But you will be at the heart of teaching and learning; and if you use good sense you will profit much and so will your pupils. Moreover, you will find, if you do your work carefully, that reports of what you have done will be gladly received by periodicals for publication. The extent to which teachers are already writing reports of a research character may be inferred from the fact that something like fifteen articles by teachers recently appeared in a single volume of one of our foremost educational journals. These articles had to do with a wide range of topics, indicating among other things that the interests of teaching include many matters outside of the classroom. The articles to which I refer dealt with extra-curricular activities, accounting systems for athletic organizations, student publications, educational opportunities for high-school graduates, the high-school assembly, the proper use of the teacher's time, tests and remedial measures, program arrangements for pupils belonging to different mental groups, and the pupil of low intelligence quotient.

It is not at all unusual for the best technical journals to contain reports of psychological experiments which public-school teachers might have conducted. For example, in one of the foremost French periodicals I lately read an interesting contribution in which the author took up one of the questions which I considered in Chapter I. "What kind of repetitions," he asks, "must be adopted in order to assure the greatest economy of time?" He then outlines a part of Lottie Steffens' study published in a German periodical in 1900—a study which could likewise have been conducted in the schoolroom as well as in the laboratory. He confirms Miss Steffens' findings to the effect that learners when left to their own devices divide their learning task into small pieces which they learn separately—in other words, that they use the part method. Another outcome of the experiment was that there were individual patterns of procedure for each learner. Although all of them used the part method, each one used it in an arrangement peculiar to himself. They had not only fallen into the part method—a method relatively uneconomical—but they had also fallen into various extremes of it, the more extreme being the least advantageous.

The point I am getting at, however, is not the outcome of the experiment, but the fact that it appropriately belongs to the school. If you substitute the teacher for the psychologist, the schoolroom for the laboratory, and a class of pupils for the half dozen learners, you have the conditions for a very much more fundamental and satisfactory experiment than the one to which I have just called your attention. Moreover, there is no reason why the report of such an experiment should not be readily accepted for publication.

Here is another experiment which was reported in a high-grade periodical and which might well have been performed by a teacher in a classroom. It has to do with speed of learning in relation to retention. There is a rather general belief that those who learn quickly forget quickly. In other words, the proverb "Easy come, easy go" is usually accepted. Norsworthy conducted an experiment in which eighty-three students learned a German-English vocabulary. Precautions were taken to make sure that no one knew any of the words before the learning took place. A record was taken of the time required by each subject to learn the vocabulary to the point where he could give all the German words when the English words were presented. Naturally this time varied considerably. The work was then put aside for thirty days without giving the students any idea that it would be referred to again. At the end of this time each student was tested on the vocabulary he had learned. Among other things, it was found that the most rapid learners remembered the greatest number of words after the thirty-day period.[1]

I have already pointed out the fact that this experiment is essentially a schoolroom experiment. The fact that it was performed by a psychologist and reported in a psychological journal does not thereby make it a problem which belongs solely to the psychologist. The question itself—the question whether rapid learning is well retained—though interesting, is completely overshadowed by a problem which these data suggest but which Norsworthy does not mention. She found and noted the great variation in rate of learning among these eighty-three students, and similar variation would undoubtedly be found by any teacher who performed the same experiment. A more important problem, however, than the relation of this variation to differences in retention would be an analysis of the methods of superior learners with a view to teaching these methods to those whose learning was poor in quality.

---

[1] Norsworthy, Naomi. "Acquisition as Related to Retention," *Journal of Educational Psychology*, 3:214-18 (April, 1912).

It is just at this point that the teacher as an experimenter has greater opportunities than the psychologist. The latter cannot ordinarily command the services of a large number of learners and even if, as in the case of Norsworthy, a large number may be secured, they cannot be secured for a long period of time. Moreover, even if the conditions were entirely favorable for the psychologist, nevertheless the problem as pursued by him is always relatively theoretical and remote. And it should be. If the story were mine to tell I could dwell with conviction upon the value of both the theory and the remoteness of some types of psychological investigation. But the time comes when the same problems should be pursued in their immediate and practical bearing. These problems in these bearings can never, in my judgment, be pursued by anyone as successfully as by the teacher. This seems to me to be true for this reason if for no other,—in the case of the teacher the application may be made to the very persons from whom the data were secured and in respect to whom, therefore, their validity is less doubtful.

Naturally when we come to a consideration of the school subjects the horizon widens. The opportunity to measure success in the learning and teaching of these subjects—an opportunity which the test movement has brought to us—enormously increases the range and variety of accessible problems. The lack of convincing means of measurement until very recent years may perhaps account for the fact that so many apparently simple problems remain unsolved.

For example, there have been advocates of oral arithmetic and written arithmetic. Most teachers use both. But no one knows which is better or whether a blending of the two is to be preferred to either. Suppose a teacher were to organize in his class two groups of children each equally capable in arithmetic. Suppose he were to teach one of them by the oral method and the other by the written method, devoting the same amount of time to each method. An appropriate work unit might be covered in this manner with a test at the end of the experimental period. No doubt few teachers, however, would wish to rely wholly upon either method. Suppose then that in a given school system we have teachers coöperating in pursuing this problem. Some teachers might devote ten minutes to oral and thirty-five minutes to written work each day. Others might give fifteen minutes to oral work and thirty minutes to written work. Still others might give twenty minutes and twenty-five minutes respectively; and further combinations might be made. All these teachers might work on a definite topic, such as interest by the six per cent method, or they might work on a series of topics. Parallel and final tests agreed upon by the group might be given at the beginning and end of the experimental period. The amount of improvement in the different groups would be a fair indication as to the best proportion of oral to written work. Later testings—

after a week, a month, or a year—might be made to see which arrangement of oral and written arithmetic produced the best retention.

Considering the vast number of teachers who have taught little children their number combinations, it is surprising to find how little they have contributed on the problems connected with this field of teaching. Should addition and subtraction facts be taught together or separately? Some teach all the addition facts before teaching any of the subtraction facts. Others teach 8 less 5 are 3, along with 5 and 3 are 8, and similarly for other pairs of addition and subtraction facts. I have talked with many teachers on this point. They disagree sharply and few of them are able to give reasons for their practice. Perhaps one way is as good as the other. It would be worth while to know this. The reason why we really know nothing about the question is because the teacher is the only one who can supply the answer—and the teacher fails to do it. Again, two equal-ability groups may be set up. One may be taught one way and the other the other way. A definite amount of teaching material may be covered, although the most convincing experiment would be one in which all the addition and subtraction facts were taught.

In some schools the subtraction facts are taught additively; in others, by the take-away method. No convincing evidence is forthcoming as to which method is the better. Again, the reason is because we have here a teacher problem; and the teachers are silent on the subject.

I could go on almost indefinitely mentioning problems in the narrow field of primary arithmetic—problems the simplicity of which might lead us to expect a ready solution. Should combinations be taught in regular order or should a difficulty order be observed? Should combinations be kept apart which have a common element, *e.g.*, $5 + 4$ and $3 + 4$? To what extent may transfer from fundamental facts to higher decade facts be relied upon? Is drill in connection with games superior to drill from a chart? How much drill should be given in the form of verbal problems? In formulating problems should we seek to make the problems interesting or merely clear?

It is by no means necessary that you should set up formal experiments involving control groups in order to serve the cause of education as a research worker. The type of investigation to which the clinical psychologist is devoted is available for the teacher. I do not mean that the teacher should use the materials or work with the same purpose that the clinician does. But I do mean that, like him, the teacher may be concerned with the study of individuals. In these studies the effort is to secure completeness not by the accumulation of *many* cases but by the extensive treatment of *one* case. The teacher as he gets experience encounters many interesting cases. If he had the habit of making records, how valuable to him and to others these cases might be. By a "case" I mean a child with his recorded characteristics, antecedents, needs, and achievements —a child who is attending school and who at various points in his school career presents problems which are met by some kind of procedure. The record of such a case while containing personal data will also report what action was taken and with what result.

A collection of such cases would be a priceless legacy from an experienced teacher. Why should new teachers be compelled to tread the same weary round and make the same mistakes year after year? Can we not save them and their pupils from some of these? For there are few new problems in teaching. There are doubtless new phases of old problems but it is scarcely likely that any really new crux ever appears in a schoolroom. The overgrown and backward child, the child who is shy, the child who has a special disability in mathematics, the sullen and the sunny, the honest and the dishonest, the leader and the follower—all these have been attending school time out of mind. They have been coaxed or cudgeled, stimulated or repressed, punished or rewarded. The inexperienced teacher must meet all these conditions as if they had never been met before. To him the experience of his predecessors is nothing. He must face the same difficulties that others have faced and fail or succeed for the same reasons that have influenced thousands of others. He has no guidance from them, no friendly lessons of caution from their mistakes.

Among the many types of research work available to teachers, the making of case studies is by no means unimportant. It involves record making and record making is not one of the teacher's strong points. It would be well if this were otherwise. At any rate I venture to believe that if the sort of record making to which I am referring could be developed in a few centers, and if a first "book of cases" could be made, such a book would justify by its usefulness every effort by which it was produced.

Let no one persuade you, then, that your opportunities for study and investigation are small. Do not suppose that educational research is appropriate only for the university professor and for those who belong to bureaus of research. The field of research has no limits other than those of education itself. Indeed research is not really a field at all. It is a method; it is a point of view. No matter how we may divide education, each subdivision has its research aspect. The kindergarten, the elementary school, the secondary school, the college, and the graduate school each has its appropriate problems of research. In colleges and universities faculty members are appointed, promoted, and paid not only because they can teach but also because they can conduct original investigation. In the high school, however, and in the elementary school, faculty members are hired to teach and not to investigate. They are not expected to concern themselves with the problems of their calling.

Yet nothing can be more certain than that if we are to make our elementary and secondary schools the schools they should be, the public-school teacher must take the field. We cannot rely upon the colleges and universities to do the work. Administrators, superintendents, supervisors, and teachers in the elementary and secondary schools should be expected to participate. Teaching and research should be required of faculty members not only in higher educational institutions but also in the public schools. The proportion of attention devoted to these two activities may indeed vary at different levels but in all schools teaching and research should be expected, recognized, planned for, and paid for.

This would be desirable even if the number of problems awaiting solution was not extremely large and even if the urgency of these problems was not immediate. It would be desirable even if the public-school people were not the very ones who could attack many of these problems with the best prospect of success. In short, it would be desirable even if no account were taken of the results as contributions to knowledge. The spirit of research among teachers would be justified merely in its reaction upon the teachers themselves. We often hear it said in university circles that the research activities of faculty members give vitality to their teaching. This is indeed true. It would be true to an even greater degree among public-school teachers, because with them the research work may be concerned not only with subject matter but also with children.

Indeed, when looked at from this point of view the teacher's occupation becomes fascinating. He has children to study—not stones, bugs, fossils or old manuscripts, but the most interesting of all possible materials—namely, human beings. Moreover, he has at hand human beings at their most engaging period—childhood and youth. And his children never grow old. In constant procession they present to him, always at the level of childhood, their innumerable interesting aspects. Yet each is different from the other—different in strength, talent, and character; different in origin, growth, and need. If teaching these children is to include studying them, the job of teaching takes on new dignity. Its scope is broadened. Its meaning is enriched. No other calling may then be compared with it. It is the great adventure.

INDEX

Ability, group testing of, 28, 29
Accounting, child, 98
Age, 146, 147, 152; mentality and, 155; school ability and mental, 160; scores, 185, 204; grade and, 187; "achievement," in reading, 205
Alexander, Jessie M., 283
Algebra, suggestions for teaching IXB, 361
Analysis, of error, 312; a method of, 342
Arithmetic, 15, 42, 45, 75, 80, 90, 171, 194-195, 315-324, 334, 342-362, 375
Army tests, 141, 154
Ashbaugh, E. J., 135
Asker, William, 254
Augur, Roman, 137
Average, computation of, 115-121
Ayres, L. P., 170, 180-184, 192-194

Bagley, W. C., 42, 164
Ballard, P. B., 250, 254
Barr, A. S., 229, 241
Barthelmess, H. M., 254
Basic combinations, 44; facts and principles in teaching, 45
Batson, W. H., 254
Bibliographies, 131, 132, 166, 197, 254-257, 293, 294, 329, 367-368
Binet, 72, 139, 142
Binet-Simon: test, 139, 140, 283; scale, 141, 149, 156, 280
Book, William F., 11, 62
Boston: schools of, 74, 75; textbook, 75; schoolmasters of, 76
Brandenburg, G. C., 231
Brinkley, S. G., 254
Brooks, *Improving Schools by Standardized Tests*, 134
Brooks, S. S., 333; graph card, 333
Buckingham, B. R., problem test, 181-184, 236, 265, 334, 361
Buckingham-Stevenson, geography test, 180, 228, 236
Burgess, W. R., reading test, 199
Burt, Cyril, citation from, 202, 203
Bush, Emma D., 290

Caldwell, Otis W., 73-75
California, University of, Southern Branch, 291
Captions, 85
Cards, guide, 94-98
Chapman, J. C., 254
Charters, W. W., 158, 244, 309, 314
Children: motor-minded and mentally deficient, 53, 54; gifted, 156, 161-163, 291; ability of, 260
Clapp, Frank L., 44, 310, 362
Classes, small and large, 54
Classification and grouping, 258-294
Coast guard service, index card, 96
Colvin, S. S., 254
Combining intelligence and educational tests, 198-209
Cooley, A. M., 254

Cornman, Oliver P., 314
Correlation coefficient, 130, 131; foot-rule method of computing, 130
Counting, errors in, 321
Counts, 314
Courtis, S. A., 73-75; tests, 140, 171, 180-184, 187, 188, 195, 196, 333, 362, 365
Cross-cut test, 232
Curriculum, 2, 3, 272, 309, 328
Curve, learning, 58-61

Dalton school "contract," 194
Data, filing statistical, 77-98, 100-102
Des Bancels, J. Larguier, citation from, 30
Detroit plan, 187-192; class record sheets, 189, 191
Deviations, plus and minus, 117, 118
Diagnostic test, 171, 172, 229, 335, 365
Dickson, V. E., 166
Discussion test, 251, 335
Division: errors in, 316; long, 317
Double-entry score sheet, 337-341
Drill or review, 38, 325
Dunn, Fanny W., 12

Ebbinghaus, H., 72, 142
Education, 1, 3, 5; Winch on, 370
Educational tests, 167-197; use of scales in, 167, 170-177; combining intelligence and, 198-209
Effect, law of, 8, 9, 12, 13, 63, 67, 70
Efficiency man, 77
Emotional side of learning, 63, 67
Engelhardt, N. L., 82, 88
Environment, 4, 6

Error: analysis of, 312-318; seriousness of, 319; prevention of, 321. *See also* Multiplication, errors in; *also* Problems, errors in; Division, errors in
Error and failure, educational meaning of, 295-329
Evaluation of tests, 177
Examination: Illinois, 175, 206, 231, 232; recall, 213; recognition, 218; "true-false" type of, 219; "short-answer," 253; "interview" type, 344. *See also* Test
Examinations: new type, 210-257; time saving in rating papers in, 247
Exercise, law of, 8, 9, 13, 16, 67

Fatigue, mental, 55-58
Fernald, Grace M., 52, 53, 70
Filer, H. A., 254
Filing, 88, 100; sample card for, 96
Fiske, T. S., 255
"Foot-rule method," 130
Forgetting, 36; how to counteract, 37
Forms, report, 83-85
Formula, for computing correlation coefficient, 130, 131
Fractions, errors in, 317
Franzen, Raymond, 206
Freeman, F. N., 170
Frequency tables, 103, 105, 107, 109-113; how to find median in, 102-132

Gates, A. I., 255
Geiger, Ruth, 54, 364
Geography: filing in relation to, 93-96; Buckingham-Stevenson test, 180, 228, 236

Geyer, Denton L., 285
Gilchrist, 65, 66
Gilliland, A. R., 197
Glass, James M., 284, 288
Godsey, Edith R.. 245
Goodrich graph cards, 333
Grade norms, 181, 182
Graded system, 258-295
Grades: percentage, 136; promotion and, 259; individual differences in, 260; segregation in, 262; grouping within, 276-281
Grammar: test methods in, 244; errors in, 314
Graph, pupil's, 184; cards, 333
Graphometer, 65, 140, 167-169
Gray, W. S., oral reading paragraphs, 140, 170
Greece, Ancient, 5
Gregory, C. A., 197
Grouping and classifying, 258-294
Group intelligence tests, 141
Growth and learning, 3

Hahn, H. H., 255
Handwriting, 167-171, 191-194; scale, 167, 180; class record sheet in, 191
Harvard-Newton scales, 171
Heck, 57, 58
Heilman, 309, 310
Hillegas, M. B., composition scale of, 140, 170; number analysis of, 360, 361
Hines, H. C., 166
Horace Mann Elementary School, report form of, 83-86

Illinois, examination, 175, 206, 231, 232; general intelligence scale, 333
Imagery, types of, 48-52
Impromptu tests, 90
Indentation, in report headings, 85
Indexing, 88, 94-98
Information question test, 228
Instruction: mass and individual, 54, 295-299, 330-368; necessity for individual, 144
Intelligence: definition of, 148; tests, 133-166, 283, 284; scales, 156, 180, 333; conservation of, 162; standards of, 201
Inventory tests, 335, 336

Jordan, A. M., 12
Jordan, R. H., 197
Jost, 18
Judgments: teacher's, 145; age effect on, 146; percentage, 134
Junior high school, 275, 281, 282

Keller, Helen, 52, 70
Keyes, Charles, 303
Kirby, Thomas J., 15, 18, 244
Knight, F. B., 255
Kuhlmann, 36

Laird, Donald A., 250, 255
Lakeman, Mary E., 33
Latin, 227, 245, 361
Laws of learning: three fundamental, 7-13
Learning curve, the, 58-61
Learning: passive and active, 4; three fundamental laws of, 7-13; logical, 20, 21; types of, 21, 22, 24, 25, 30, 31, 33-35, 48-54; measuring, 25-27, 68; emotional, 63, 67; trial and error, 311. *See also* Laws of learning
Lippincott-Chapman test, 175

Long division, errors in, 317
Lowell, group intelligence scale, 156
Lyon, D. O., 37, 38

McCall, W. A., 197, 250, 253, 255, 265, 341
McKinney, 302
Mann, Horace, 73, 74, 76, 258; Elementary School, 84
Mass teaching, 295
May, M. A., 255
Mean or average, 103-132
Median, the, 102-115; frequency tables relating to, 103, 105, 107, 109-113
Memory, 25, 26, 29, 30, 31, 69, 72, 88; retention, 373, 374
Mental fatigue, 55-58, 70
Mentality and age, 155
Meumann, E., 4, 23, 24, 34
Milk, measuring cleanliness of, 168, 169
Milton, John, 214
"Mixed" method of learning, 33, 34, 68
Monroe, W. S., tests, 171, 180, 181, 185, 197, 205, 342, 344
Morrison, J. Cayce, 39
Morton, 314
Motor learning, 52, 53
Moyer, Edward L., 286
Multiplication, errors in, 316
Muscular sense, 48
Myers, Garry C., 314, 319-321, 324

Name, in report forms, 79, 84
Neatness: transfer of training in, 42; teaching of, 46
New-type: examinations, 210-257
Norms, 146; grade, 181; table of norms on certain tests, 182, 183
Norsworthy, Naomi, 373, 374
Number combinations, 44, 376, 377

Observational learning, 19-22
Odell, C. W., 131
Ohio: State University, Department of Agricultural Engineering test, 236-239; State Department of Education, Classification Committee, 265, 268, 269, 274
Opposites test, 231
Oral reading, 140
O'Rourke, L. J., 254
Osburn, W. J., 45, 171, 314, 315, 317
Otis, Arthur S., 131, 141, 232-234

Papers, filing test, 90
"Part" and "whole" methods of learning, 22-25, 29-35, 68
Pechstein, 24
Percentage: judgments, 134; grades, 134, 136-139
Pintner, R., 156, 197
Plateaus in learning curve, 61-63
Plato, 6; *Republic* of, 5
Practice, periods of, 13-19, 68
Précis writing, 82
Pressey, L. C., 156, 175, 197, 283, 284
Pressey, S. L., 156, 197, 283, 284
Probation, 304-308
Problem tests, 181, 345-360
Problems, errors in, 318, 345-360
Progress: need of more exact knowledge with, 76, 77; records, 91-93; books, 92
Promotion, 299-308, 327
Pupils: grouping and classifying, 258-294; ability of, 260; capable, 345; individual differences in, 346

Purdue English test, 224, 231
Pyle, W. H., 30, 31

Rapid learners, 69
Rating: unreliability of teacher's, 134-139; effect of teacher's, 138; time saving in, 247
Readiness, law of, 7, 9-11
Reading, 10-12, 21, 43, 50, 52-54; silent, 50, 180-184; tests, 178, 185, 199, 232, 364, 365; age, 205; primary, 232; scale, 342; problem solving and, 350
Recall examination, 213; completion, form of, 215
Recognition tests, 213, 217
Record forms, 77-82, 86, 87
Records, classroom, 98, 99
Recreation and play, 57
Register, school, 98, 99
Regrading, 263-269; by subjects, 274, 275, 276
Reminiscence, 39, 69
Repetition, 19, 20, 21, 67, 68, 326
Report forms, 83-87; time limit in, 83, 84
Republic, Plato's, 5, 6
Research work, opportunity in, 16, 71, 72, 293
Retention, 25, 26, 32, 33
Rhythm, in learning, 49
Rice, J. M., 198
Riegel, Vernon M., 265
Ruch, G. M., 250
Ruediger, 42
Rugg, H. O., 132

Scales: Ayres, 109, 170, 180-182, 192-194; Hillegas, 140, 170; Binet-Simon, 141, 156, 280; Thorndike, 169, 170; Harvard-Newton, 171; Willing, 171, 182; Illinois General Intelligence, 180; Stone, 342
School, and society, 3; old and new, 73-77; improvement of, 134; junior high, 275, 281, 282
Scores, 180-183; age, 185
Score-sheet, 80, 81; double entry, 337-341; sample of, 340
Shakespeare, 298
"Short-answer" examinations, 253
Silent reading, 50, 180-181; Courtis test in, 184
Similarities test, 233
Snyder, J. C., 30
Spelling, 75, 109, 110; Detroit plan of rating, 187-192, 309, 336
Stanford achievement test, 175, 206, 216
Starch, Daniel, 14, 15, 17, 18
Statistical clerk, 86
Statistician, 77
Statistics: needed by modern teacher, 73-132; record forms, 77; items and sub-items in, 87; of variation, 121-126; of relationship, 126-132
Stedman, Lulu M., 291
Steffens, Lottie, 372
"Step" in frequency table, 103
Stern, 142
Stetson, Paul C., 288
Stevenson, P. R., 165, 236, 240
Stone narrative reading scale, 342
Strayer, G. D., 82, 88
Sub-items: position of, in reports, 86, 87
Subject-matter tests, 139-141
Subtraction, errors in, 316
Survey test, 330-333
Synonyms test, 231
System, graded, 258-295

Tables, relating to median, 111-113

Teacher, 1, 7, 12, 39, 45, 47; advantages over psychologist, 71, 72, 77-79, 81, 82, 88; professional growth of, 89; in planning work, 91; progress books as aids to, 92, 134-136; as research worker, 369-380; opportunities for, 374-378. *See also* Judgments

Teachers College, Columbia University, tests, 170, 196

Teaching: factors to be considered in, 45, 88, 89; mass and individual, 295; emphasis in, 309; speed in, 325; application of error to, 328

Terman, L. M., 154, 155, 156, 166, 231, 236

Test: papers, 42, 90, 91, 95, 343, 344, 366, 543; standardized, 80; impromptu, 90; intelligence, 133-166; educational, 167-197; problem, 181-183; subject-matter, 199, 200; recall type, 213, 214; completion, 215-218; alternative-response, 218-224; true-false, 219-221; multiple-choice, 224-229; information-question, 228; thought-question, 229; synonyms, 231; opposites, 231; cross-out, 232; similarities, 233; analogies, 234; best reason, 235; pictures, 236; matching, 239; definitions, 240; cause and effect, 241; combinations, 242; diagnostic, 334; inventory, 335

Tests: simplification of, 172-173; manufacture and sale of, 176; critical evaluation, 177; combining intelligence and educational, 198-209; new-type examination, 210-257; how to make new-type, 246

Thorndike, E. L., 12, 63, 65, 140, 156, 167, 169, 180, 193, 312, 313, 315, 324

Thought-question test, 229

Thurstone, L. L., 132, 156

Torgerson, T. L., 286, 287

Trabue, M. R., 66, 179, 197

Transfer of training, 40-47, 69

Transmuting, test scores, 180

True-false tests, 219-221, 250, 338, 339

Types, of classification, 262; of learning, 20, 48-54, 70; of error, 312-314, 318

Uhl, W. L., 314

University of California, Southern Branch, training school of, 291

Van Denburg, J. K., 281, 282

Van Scoyck-Young Permanent Test Record Cards, 332, 333

Variation, statistics of, 121

Varner, G. F., 147, 148

Visual imagery, 48, 49, 51

Washburne, Carleton W., citation from, 362, 363

Watt, Henry J., 31

Welch, Emily H., 92

"Whole" and "part" methods of learning, 22-25, 29-35, 68

Willing scale, 171; tests, 182, 183

Wilson, Estaline, 351

Winch, W. H., 24, 25, 42, 43, 57, 370

Winnetka, schools at, 362, 363

Woodrow, Herbert, 154, 155, 156, 166

Wundt, 142

Yates, 161

SCHARGED

AY 18 '74

SCHARGED

DISCHARGED

SCHARGED

LB
1027
.B8

20518

STATE TEACHERS COLLEGE

Stevens Point, Wis.

Books may be kept two weeks and may be renewed for one week, except reserved books and magazines. These may be drawn only for overnight use.

A fine of five cents a day will be charged for reserved books not returned on time, and a fine of two cents a day for all other books. No book will be loaned to anyone owing a fine.

LB1027.B8
Research for teachers
3 1775 [redacted]